Statistical Modelling of The Consistency of Symptoms Reported During Hypoglycaemia for Individual Patients

Hani Syahida Zulkafli

Submitted for the degree of
Doctor of Philosophy

Heriot-Watt University

Department of Actuarial Mathematics and Statistics,
School of Mathematical and Computer Sciences.

May 19, 2017

Abstract

In this thesis, we use Bayesian methodology and Markov chain Monte Carlo techniques to construct logistic-type latent variable statistical models for estimating the consistency of hypoglycaemic symptoms experienced by individual diabetic patients. Consistency in reporting experienced symptoms of hypoglycaemia is related to early detection of symptoms and is therefore important for fast corrective action. Based on a model developed by Zammit et al. (2011) we classify symptoms into different groups and consider between-groups variability. Our work also explores a number of possible symptom-experiencing thresholds that can be used in the consistency model. To evaluate the performance of each consistency model, we develop ideas based on Bayesian latent residuals (Streftaris and Gibson, 2012) to check on the models' fit and utilise posterior predictive checking methodology (Gelman et al., 1996 and Streftaris et al., 2013) to assess relevant performance. The impact of using data from hypoglycaemic episodes occurring within 24 hours from an earlier episode is also explored using various approaches, as previous work claims that such episodes might lead to diminished intensity of the episodes. Using generalised linear-type model methodology, we investigate how various factors such as age, gender, type and duration of diabetes, body mass index, retinopathy and others, or their interaction, can affect patients' consistency. Additionally, we develop a hierarchical model that is able to estimate consistency and identify factors affecting it in a single setting. Finally, we work on determining the best sets of variables for a predictive model. For this purpose, we use Gibbs variable selection and a stepwise regression procedure. Due to model uncertainty, we apply Bayesian model averaging to a number of selected models given by Gibbs variable selection.

Acknowledgements

First and foremost I would like to express praise to almighty Allah with His compassion and mercifulness to allow me finishing this PhD research. Finishing this research means a lot to me and could not have been possible without the help of many people.

I would like to express my sincerest appreciation and gratitude to my supervisor, Dr. George Streftaris for his patience in guiding me, constantly providing me with valuable critiques, recommendations and intellectual comments. I really appreciate being available for regular meetings which helped me gained valuable knowledge. Thank you for always lending time for checking this thesis.

I also thank my second supervisor, Professor Gavin Gibson for his continuous support and making important suggestions for this research.

I would like to thank Dr. Nicola Zammitt for sharing her expertise especially in medical aspects.

I am indebted to Iain, Claire, June, and Christine for their help and technical support during my study in Heriot-Watt University.

I thank the Government of Malaysia and Universiti Putra Malaysia for providing the financial support and opportunity for my PhD degree.

Finally, I owe deep sense of gratitude to my family for all the encouragement and faith. They have been exceptionally supportive throughout.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Hypoglycaemia

Diabetes is a health condition caused when glucose level in blood is high. Beta cells in the pancreas produce insulin hormone which helps unlock cells for glucose to enter. When the blood glucose level gets high, insulin is secreted and aids glucose to enter cells, thus lowering the glucose level back to normal. Without insulin glucose will accumulate in the bloodstream, raising the blood glucose level. Glucose needs to enter cells before it can be used as energy source for our body to do daily activities. In short, glucose acts as fuel to our body to live life.

The two most common types of diabetes are Type 1 and diabetes Type 2 (National Diabetes Data Group, 1979). Type 1 diabetes is a condition caused when the body's immune system mistakenly destroys beta cells as if they are a harmful pathogen (Atkinson and Maclaren, 1994). As a result, insulin cannot be produced and glucose cannot enter cells from bloodstream. Type 1 diabetes is also known as 'Juvenile Diabetes' because it is usually diagnosed in young people. Nonetheless, it is not rare for adults to develop Type 1 diabetes nowadays. This type of diabetes is called insulin dependent diabetes because patients need to rely on injected or pumped insulin to help maintain their blood glucose level. Diabetes Type 2 is a condition when either not enough insulin produced due to damaged beta cells, or insulin is there but not functioning properly in helping cells take in glucose (Reaven et al., 1976 and Olefsky et al., 1982). Type 2 diabetes has higher prevalence than other types. 90% of diabetic

cases worldwide are Type 2 diabetes. Some of the factors that may increase the risk of developing this type of diabetes are being overweight (Kolterman et al., 1980 and Bogardus et al., 1985), lack of exercise, unhealthy diet, smoking and history of first degree relatives with Type 2 diabetes.

Hypoglycaemia is a problem of glucose deprivation in the brain that is the result of the blood glucose level in the body becoming abnormally low, usually bellow 4mmol/L. This condition is referred to as insulin reaction because it is caused by insulin-treatment in diabetic patients. Glucose is the body main source of energy. When carbohydrate is broken down into glucose, it is directly absorbed into the bloodstream. However, glucose cannot enter a cell without insulin; a hormone secreted by pancreas that acts as a key to allow glucose to enter cells. For instance, without insulin a cell will be starved of energy.

Diabetic patients, either have malfunction of insulin (Type 2), or no insulin at all (Type 1), and rely on insulin treatment to help maintain their blood glucose levels. However, too much insulin will drop the glucose level in bloodstream and trigger hypoglycaemia. There are three possible conditions that can cause excess insulin in body;

- Too much insulin released in bloodstream (e.g. increased insulin dosage, Banarer and Cryer, 2004).
- Body's sugar is used up too quickly (e.g. intense exercise, Banarer and Cryer, 2004).
- Glucose is released into the bloodstream slowly (e.g. eat later than usual or eat smaller portion than usual).

In the second and third situations, even a normal dose of insulin can be too much because the insulin required by the body at that moment is lower than usual. Not having meals at the usual time or irregular meals are the most common factor that lead to hypoglycaemia (Nattrass and Lauritzen, 2000 and Sotiropoulos et al., 2005).

An event of hypoglycaemia occurrence is referred to as an 'episode' throughout this thesis. Hypoglycaemia episodes can range from mild, which the patients can treat themselves, to severe which require other people's aid (American Diabetes Association, 2005 and Canadian Diabetes Association, 2003). They may result in a medical emergency when a person becomes unconscious during a severe hypoglycaemia episode. A doctor or a family member needs to inject glucagon hormone in an attempt to bring the glucose level back to normal. This hormone helps to simulate liver to release stored glucose when glucose level is too low.

It is very important for patients to be able to detect the onset of hypoglycaemia so that they can take necessary action to bring their declining glucose level back to normal. Failure to detect this will prevent immediate corrective action resulting in a severe hypoglycaemia episode, where patient needs other people's help to recover. Untreated hypoglycaemia may impact on brain power (because brain needs glucose to function), cause cognitive decline or dead-in-bed syndrome. Longer duration of diabetes and insulin treatment increase the risk of hypoglycaemia (Henderson, 2003).

A symptom is defined as any subjective indication of disease that is apparent to the patient. All symptoms of hypoglycaemia are common symptoms that can sometimes be caused by another health problem or illness, rather than being exclusive to lack of glucose. Therefore the only way to confirm an episode is by measuring the blood glucose level using a glucose meter. There is a wide range of symptoms of hypoglycaemia. By analysing insulin-treated diabetic patients using factor analysis, Hepburn et al. (1993) concluded a list of 11 key hypoglycaemic symptoms; sweating, pounding heart, shivering, hunger, confusion, drowsiness, difficulty speaking, unsteadiness, nausea, headache and odd behaviour. These 11 symptoms were identified as the Edinburgh Hypoglycaemia Scale. These symptoms vary between individuals (McAulay et al., 2001, Zammit et al., 2011) and are idiosyncratic (McAulay et al., 2001). There is variability of symptoms experienced between people (Cox et al., 1993).

A recent study by Zammit et al. (2011) revealed that symptoms of hypoglycaemia of an adult patient also vary across episodes. Variability both within and between

individual patients makes it difficult for patients to precisely detect the onset of hypoglycaemia episodes, with between-variability hindering the precise identification of a condition-specific set of symptoms. Thus, it is crucial for them to be able to identify a sufficient number of symptoms that are typical for themselves and understand that they may experience different sets of symptoms in different episodes. Patients with four or more reliable symptoms manage to detect their low glucose levels correctly in 75% of the hypoglycaemic episodes, while patients who have fewer than four reliable symptoms, only manage to correctly recognise their low sugar levels half of the occasions (Cox et al., 1993).

1.1.1 Earlier classification of symptoms

Symptoms of hypoglycaemia can be classified into groups based on their cause. Symptoms that develop from lack of glucose in the brain are classified as neuroglycopenic symptoms whereas symptoms that happen unconsciously resulting from the nervous system's response to hypoglycaemia are categorised as autonomic symptoms. Hepburn et al. (1991) assessed the classification of symptoms using factor analysis and managed to separate the symptoms into two distinct groups. They allocated symptoms dizziness, confusion, tiredness difficulty speaking, headache and inability to concentrate into the neuroglycopenic group, and symptoms sweating, trembling, warmness, anxiety and nausea into the autonomic group. Analysis from the same researchers (Hepburn et al., 1992) on a larger sample of insulin-treated patients derived five different groups of hypoglycaemic symptoms; neuroglycopenic, autonomic, general malaise, neurological dysfunction (motor) and neurological dysfunction (sensory).

Deary et al. (1993) came up with three groups that best partitioned the symptoms. They suggested that confusion, drowsiness, odd behaviour, difficulty speaking, unsteady and headache belong to the neuroglycopenic group. The second group is the autonomic group, which includes symptoms sweating, shaking, hunger and pounding heart. Symptoms nausea and headache were classified under a third group characterised as the " general malaise" group.

Two of the commonly identified hypoglycaemia symptoms are debatable as to which category they belong to. Although the "feeling warm" symptom appeared as an autonomic symptom in Hepburn et al. (1991), it was not prevented by drugs that block the autonomic nervous system in physiological studies (Towler, 1993). The same goes for the "irritability" symptom. Nervousness and anxiety are autonomic but altered behaviour is neuroglycopenic. So they can be autonomic, neuroglycopenic or both depending on the interpretation of the patient filling in the questionnaire.

1.2 Aim of the thesis

In this thesis, our ultimate aim is to develop, fit and assess statistical models for the consistency of symptoms experienced by individual patients during hypoglycaemia using Bayesian methodology and Markov Chain Monte Carlo (MCMC) techniques for estimation. We calibrate a consistency estimation model developed by Zammitt et al. (2011) to consider between-group variability when symptoms of hypoglycaemia are classified into groups and also develop a hierarchical model that can estimate patients' consistency and identify factors affecting it in one single setting. We also extend previous work to investigate the impact of interactions among patient-specific factors on symptom consistency and then determine a relevant predictive model. We employ Gibbs Variable Selection with different prior distribution settings (Dellaportas et al., 2002, Ntzoufras, 2002, 2011, and Ntzoufras et al., 2003) and stepwise regression in order to choose the best set of covariates for our predictive model. One of the main concerns when trying to obtain the best predictive model is model uncertainty. We tackle this problem by using Bayesian Model Averaging. With several consistency models developed, our challenge is to evaluate the performance of each model before making decisions on which model can give better consistency estimates. Therefore, we develop ideas based on Bayesian latent residuals (Streftaris and Gibson, 2012) to check on the models' fit and we also perform assessment based on posterior predictive checking methodology (Gelman et al., 1996 and Streftaris et al., 2013).

The data used in our analysis were provided by the UK Hypoglycaemia Study Group and will be discussed further in Chapter 2. We initially received data on hypogly-

caemia episodes from 59 patients who experienced at least 19 episodes in a period of 9-12 months. At a later stage we received additional data concerning seven patients with fewer than 19 episodes to include in our work. Some of the episodes occurred within 24 hours from their preceding episode. Thus, we also explore whether there is any significant difference or association between the intensity of these episodes and other episodes.

All simulations to obtain parameter estimations in this thesis are mainly carried out in WinBUGS (Spiegelhalter et al., 2003). Other computational work is done with the help of the R software (R Core Team, 2013). This is a software suite that can be used for Bayesian analysis employing MCMC methodology (see for example, Raab, 2001).

1.3 Thesis outline

The data used in this thesis are described in detail in Chapter 2 with brief explanations on some related medical aspects. Chapter 2 also provides a review of the basic symptom consistency model, which is the core model leading to further extensions and analysis in the thesis. An overview of the methodology utilised in this thesis is also given in this chapter.

We start to model individual patients' consistency using Bayesian methodology and MCMC techniques in Chapter 3. The earlier part of this chapter focuses on exploring different possible thresholds that can be used in the consistency model and then comparing their model fit using Deviance Information Criterion (DIC). Chapter 3 also discusses the expansion of the core model from Chapter 2, by adding symptoms' grouping. The final sections of this chapter explain the model assessment and model verification involving all discussed models. The methods used here involve stochastic latent residuals and posterior predictive checking.

Chapter 4 analyses the effect of adding hypoglycaemic episodes that have occurred within 24 hours from a preceding episode. This is done by analysing the correlations, relative means and coefficient of variation of the episodes' intensity. In the last section

we provide results and posterior consistency estimates when we take into account all episodes.

A study of potential factors affecting consistency is presented in Chapter 5. Generalised linear model methodology is used to model the effect of ten covariates and their interactions on symptom reporting of an individual patient. We also propose a hierarchical model which can be used to estimate consistency and investigate what covariates affect it.

We perform model selection in Chapter 6 to determine the most suitable predictive model for consistency prediction. Two methods used are the stepwise selection and Gibbs variable selection with three prior settings. Due to model uncertainty, we also use Bayesian model averaging.

In Chapter 7, we present our conclusions, discussions and ideas for further research.

Chapter 2

Data and Core Model

This chapter discusses in more detail the data and modelling used in this thesis. We first explain how the data was collected and then provide information on several medical terms and conditions related to the collected data on the date. We also present the basic model and framework for estimating individual patients' consistency, which is later expanded for other consistency models in subsequent chapters. This chapter also provides a brief review of Bayesian inference and Markov chain Monte Carlo techniques which are the core methodologies that will be used throughout this thesis.

2.1 Data

To achieve our aim of modelling the consistency of symptoms experienced by individual patients, we use symptoms reported by diabetic patients during hypoglycaemia episodes. This data was collected by the UK Hypoglycaemia Study Group in 2002-2004. The UK Hypoglycaemia Study Group initially recruited 436 subjects but 35 of them were not eligible for the study. 383 of them completed phase 1 of the study and 84% of the subjects successfully completed phase 2 of the study. Every subject that participated in this survey was given a form (see Appendix A) to record each of their hypoglycaemia episode, and was asked to return the forms on a monthly basis for 12 months. They would receive a follow-up call if they failed to do so. Note that the available data was not collected so that it was representative of the population of diabetic patients. Hence, we do not claim that the results and findings from the

analyses in the current thesis can necessarily be generalised to an entire population of diabetic patients with hypoglycaemic episodes. While our analyses do not form part of a properly randomised study, they are valid for patients with the general characteristics of the groups studied here, and we anticipate that our results can guide and inform relevant conclusions.

Besides blood glucose level, the forms collected data of date, time and duration of the hypoglycaemic episodes, treatments received during the episodes, and symptoms experienced. Some of the symptoms are impaired concentration, where the patients cannot concentrate or are unable to maintain focus on what they are doing, and double vision or medically known as diplopia. This is a condition where the patients see two images of everything. Another typical symptom of hypoglycaemia is anxiety, which is a feeling of fear, worry and being tense for no apparent reason. Patients may also experience pounding heart, which means rapid and irregular heartbeats, or sometimes feel anger and be easily upset, i.e. irritability during their episodes. Table 2.1 gives the standard list of symptoms included in the work for this thesis. Category "none" was used in the analysis to account for cases where patients recorded an episode of hypoglycaemia as a result of self-checking of their blood glucose, even if no other apparent warning symptom was experienced. This is different from the "nonspecific awareness" category, where patients were allowed to add free text.

The form also required the respondents to report if they needed assistance or lost consciousness during the hypoglycaemia episodes, as any episode that involves either one of these is considered as a severe episode.

Subjects were asked to monitor their blood glucose using a Medisense G glucose meter (Abbott Laboratories, Abbott Park, IL). Episodes with capillary glucose reading $<$ 3.0 mmol/L ($<$ 54 mg/dL) or when there was no blood glucose measurement available and the symptoms resolved on taking carbohydrate are considered valid, while episodes with capillary glucose $>$ 4.0 mmol/L are not considered as valid hypoglycaemic episodes (Zammitt et al., 2011).

Symptom	Description	Category	% of episodes
1	Confusion	Neuroglycopenic	15.76
2	Sweating	Autonomic	24.18
3	Drowsiness	Neuroglycopenic	5.77
4	Weakness	Neuroglycopenic	27.52
5	Dizziness	Neuroglycopenic	10.18
6	Feeling warm	Autonomic/Neuroglycopenic	16.23
7	Difficulty speaking	Neuroglycopenic	3.91
8	Pounding heart	Autonomic	8.74
9	Impaired concentration	Neuroglycopenic	26.84
10	Shivering	Autonomic	2.22
11	Unsteady	Neuroglycopenic	14.19
12	Nonspecific awareness	Other	4.80
13	Double vision	Neuroglycopenic	0.64
14	Blurred vision	Neuroglycopenic	12.47
15	Hunger	Autonomic	20.85
16	Thirst	Autonomic	3.83
17	Nausea	General malaise	4.94
18	Anxiety	Autonomic	13.54
19	Tiredness	Neuroglycopenic	15.87
20	Tingling	Autonomic	14.12
21	Trembling	Autonomic	20.06
22	Headache	General malaise	8.24
23	Malaise	General malaise	0.57
24	Irritability	Autonomic/Neuroglycopenic	0.93
25	Other	Other	1.47
26	None	No symptom	10.64

Table 2.1: List of symptoms on patients' report forms.

The initial part of the analysis in this thesis excludes data arising from successive episodes occurring within a period of 24 hours, to avoid issues with potential diminished intensity for episodes occurring in close time proximity, as some anecdotal medical experience may suggest. We also initially restrict our analysis to patients with 19 or more episodes following earlier work. Hence, we end up with 59 subjects who reported a total of 3,474 episodes of hypoglycaemia. These subjects consist of

cases of Type 1 (77%) and Type 2 diabetes, with age ranging from 22 to 74 years old. There is a good representation of males and females, with 65% of them being male.

Data on seven additional subjects were provided during the course of this work and were subsequently included in the analysis. Each of these subjects had reported $<$ 19 hypoglycaemic episodes. All seven patients had Type 2 diabetes. This data was included in our analysis from Chapter 3 onwards.

The study aimed to recruit 120 Type 2 diabetic respondents and 60 Type 1 respondents. Instead of selecting patients of Type 1 diabetes with shorter period, before the start of the study a pragmatic decision was made to select patients with Type 1 diabetes for < 5 years because of the need to recruit sufficient number of respondents. A summary of the data for all subjects according to their group is given in Table 2.3. We will make use of part of the information in this categorisation later, in selecting covariates that will be used to predict the subjects' symptoms reporting consistency. However, we do not take into account the type of treatment received.

Category	Description	No. of cases
1	Type 2 diabetes treated with a sufonyl-urea (tablet)	4
2	Type 2 diabetes treated with insulin for < 2 years	8
3	Type 2 diabetes treated with insulin for > 5 years	9
4	Type 1 diabetes of < 5 years in duration	21
5	Type 1 diabetes of > 15 years in duration	24

Table 2.2: Groups of patients.

For each of the participant, we were also provided with certain characteristics such as follows:

i. Gender: male, female (coded as 0 and 1, respectively)

ii. Duration of diagnosed diabetes: in months

iii. Body mass index (BMI)

iv. Age: in months

v. Type of diabetes: Type 1, Type 2 (coded as 0 and 1, respectively)

vi. Awareness of hypoglycaemia: score from 1 to 7 with higher score corresponding to diminishing awareness of hypoglycaemia

vii. Type of retinopathy: no retinopathy, background retinopathy, proliferative retinopathy (coded as 0, 1 and 2, respectively)

viii. Levels of C-peptide: in nmol/L

ix. Percentage of haemoglobin A1c

x. Levels of serum angiotensin converting enzyme: in IU/L.

This information acts as useful variables and is used in this work to investigate which of these characteristics affect the individual's consistency of symptom reporting in Chapter 5 and to determine the best predictive model in Chapter 6. In what follows we provide some information on the characteristics that may not be standard, and their association to diabetes. A list of characteristics of each patient is supplied in Appendix B.

Retinopathy is a diabetic eye disease associated with poor control of blood sugar level. In this study, we divided this condition into three categories; no retinopathy, background retinopathy and proliferative retinopathy. Background retinopathy is the mildest stage of retinopathy. At this stage, blood vessels in retina swell. Although this will not cause any eye sight problem, it has to be monitored closely or it can leads to a more severe stage. Proliferative retinopathy refers to a more serious stage where there is blockage and bleeding in blood vessels in the retina thus forming new abnormal blood vessels. This situation affects the eye-sight and has the potential to cause blindness.

C-peptide is a protein released together with insulin by the pancreas. They are released with ratio 1:1, thus making it a perfect marker for insulin production. High level of C-peptide means high insulin is produced and vice versa. Insulin helps to regulate the glucose level in blood by allowing glucose to enter cells. Excessive insulin released will results in low blood glucose level (hypoglycaemia). Therefore, C-peptide test is useful in hypoglycaemia cases, when we want to measure if a patient's body is

producing too much insulin.

Serum angiotensin converting enzyme is a protein related to blood pressure control and other functions. A person with diabetes has higher than normal level of this protein.

Glucose in blood combines with haemoglobin making it 'glycated'. This is termed as haemoglobin A1c . We can know our blood glucose level by measuring the haemoglobin A1c readings. The higher the concentration of glucose in blood, the higher the haemoglobin A1c reading. Blood glucose levels reading obtained from usual blood glucose test (i.e.: fasting glucose test), only provide indication of glucose concentration in that particular time when we do the test. The haemoglobin A1c readings on the other hand, give an average level of blood glucose in our body in a period of 2-3 months.

Presence of retinopathy is detected by retinal screening. It is a procedure which allows specialists to assessed any evidence or severity of retinopathy by taking photographs of the retina. The damaged capillaries and spots of blood in retina are generally visible as dots and blots in the photographs. Measurements of haemoglobin A1c and serum angiotensin converting enzyme were done in a central laboratory.

The identity of each subject in the study is confidential. Therefore, a unique four-digit code is assigned to each of them. There are missing values in the data set. Subject 3065 has unspecified retinopathy record, whereas subjects 1028, 1039, 2021, 3050, 4023, and 6019 have no record of the haemoglobin A1c readings. The C-peptide record for subject 4034 is also missing. The methodology for treating these missing values is discussed further in Chapter 3.

From the detailed data included in the provided forms, our analysis uses the following information:

- the reporting of symptoms in the occurrence of each episode used to model individual consistency,

- the date and time of the occurrence of each episode used to distinguish the episodes occurring within 24 hours, and
- the patients' characteristics; i.e. gender, duration of diabetes, BMI, age, type of diabetes, awareness of hypogylcaemia, type of retinopathy, levels of C-peptide, percentage of haemoglobin A1c, and levels of serum angiotensin converting enzyme to be used in investigating the effect of these factors on individual consistency.

	T2*tabs*		T2*Ins* < *2*		T2*Ins* > *5*	T1*Ins* < *5*	T1*Ins* > *15*	Total
	<19 ep.	>19 ep.	<19 ep.	>19 ep.	>19 ep.	>19 episodes	>19 episodes	
Number of subjects	3	1	4	4	9	21	24	66
Episodes per group	39	25	53	113	476	1385	1475	3566
Episodes per group after episodes < 24 h of each other excluded	39	25	52	104	370	1095	1104	2789
Number of episodes per patient	12 (10-17)	25 (25)	12.5 (11-17)	28.5 (20-36)	37 (27-146)	49 (19-210)	44.5 (19-300)	42 (19-300)
Episodes per patients after episodes < 24 h of each other excluded	12 (10-17)	25 (25)	12 (11-17)	25 (20-34)	31 (25-105)	44 (19-134)	37 (19-138)	34.5 (10-136)
Asymptomatic episodes per group (%) after episodes < 24 h of each other excluded	0%	36%	3.8%	0%	9.3%	0.9%	4.5%	11%
Number (%) male	3 (100%)	1 (100%)	2 (50%)	4 (100%)	8 (89%)	14 (67%)	11 (46%)	43 (65%)
Age (years)	64 (62-69)	51 (51)	60 (43-71)	65 (60-74)	65(57-72)	39 (22-70)	58 (34-72)	57.7 (22-74)
Number (%) with impaired awareness	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (22%)	7 (33%)	13 (54%)	22 (37%)
Body mass index (kg/m^2)	25.7 (25.1-37)	23.7 (23.7)	27.7(27.3-37.5)	27.8 (26-30.2)	27 (21.9-33)	24 (19.5-29.6)	25.3 (21.6 -42.7)	25.0 (19.5-42.7)
C-peptide (nmol/L)	1.21 (0.97-1.31)	2.22 (2.22)	0.75 (0.33-1.12)	0.85 (0.27-1.58)	0.24(0.05-0.21)	0.45 (0.06-0.87)	0.09 (0.05-0.85)	26 (0.05-2.51)
Haemoglobin A1c (%)	7.2 (6.3-7.5)	7.1 (7.1)	6.6 (6.4-9.5)	8.3 (7.8-8.8)	7.6 (6.3-8.9)	7.2 (5.6-10.1)	7.8 (6.1-9.7)	7.55 (5.6-10.1)
ACE (IU/L)	12 (7-46)	20 (20)	6.6 (6.4-9.5)	13.5 (7-24)	39 (4-71)	34 (18-94)	31.5 (3-98)	32.5 (3-98)
Number (%) with:								
No retinopathy	3 (100%)	0 (0%)	4 (100%)	4 (100%)	3 (37.5%)	20 (95.24%)	5 (20.83%)	39 (60%)
Background retinopathy	0 (0%)	1 (100%)	0 (0%)	0 (0%)	2 (25%)	1 (4.76%)	14 (58.34%)	18 (27.7%)
Proliferative retinopathy	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (37.5%)	0 (0%)	5 (20.83%)	8 (13.3%)

Data are given as median (range) unless otherwise stated. 'ep.' refers to hypoglycaemic episode.
ACE, angiotensin converting enzyme; T1Ins <5, Type 1 diabetes with <5 years' duration; T1Ins >15, Type 1 diabetes with >15 years' duration; T2tabs, Type 2 diabetes treated with oral agents; T2Ins <2, Type 2 diabetes treated with insulin for <2 years; T2Ins >5, Type 2 diabetes treated with insulin for >5 years.

Table 2.3: Subject characteristics and hypoglycaemic episodes within each group.

2.2 Data presentation

To assess consistent reporting across episodes for each subject, we present the symptoms and episodes of each patient in a matrix form with $J \times K$ dimension; J = number of symptoms (=26 for all patients) and K_i = number of episodes for patient i, see for example Figure 2.1 for Subject 6010 with K =19 episodes. In Figure 2.1(b) from the top-left corner, symptoms and episodes are arranged in descending order, from more frequent to less frequent symptom and from more intense to less intense episode. The frequency of a symptom reported by a patient represents the symptom's propensity and the number of symptoms in one episode represents the intensity of that episode.

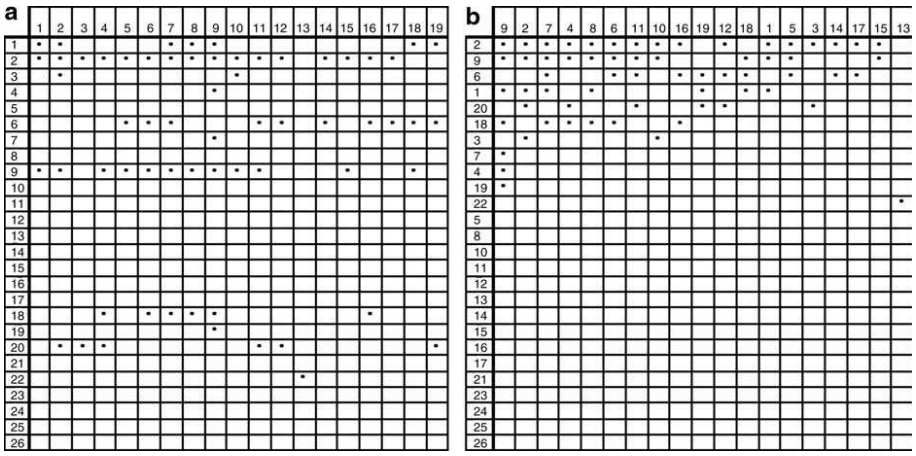


Figure 2.1: a) Example of a $J \times K$ matrix of indicator variable (J=number of symptoms; K=number of episodes) for subject 6010 with symptoms 1-26 listed vertically and hypoglycaemic episodes listed horizontally. Each reported symptom is marked with a square. b) Rearrangement of the matrix rows and columns so that rows now appear according to frequency with which symptoms are experienced and columns according to the number of symptoms per episode (both following a descending order from the top-left corner).

We assume that consistency of symptom reporting follows the principle that if a patient is consistent in reporting symptoms experienced throughout the hypoglycaemic episodes, higher ranked symptoms must be reported first before reporting lower ranked symptoms. For example, if a symptom ranked at number 7 (i.e. symptom 3 in 7^{th} row in Figure 2.1(b)) is reported at episode k then all symptoms in rows above and columns to the left must also have been reported. When this principle is violated, we will have embedded empty cells, meaning that there is at least one filled cell to the right or below the empty cell, hence suggesting inconsistency. Thus we say that

symptoms are experienced based on a hierarchical ordering determined by their latent characteristics, i.e. propensity of symptoms and intensity of episodes.

2.3 Bayesian inference

This section gives a brief review of Bayesian estimation, and an introduction to MCMC approaches (Metropolis-Hastings and Gibbs Sampling) will also be provided in the following sections.

Bayesian methodology is employed to estimate the posterior distribution of model parameters. The heart of Bayesian modelling is Bayes theorem. Suppose we have observed data $\boldsymbol{x} = \{x_1, ..., x_n\}$ from a distribution involving an unknown parameter vector $\boldsymbol{\theta}$, for which we wish to make inference. This parameter $\boldsymbol{\theta}$ is assumed to be random rather than fixed like in frequentist approach. We assume here that θ is a single continuous parameter. Then Bayes Theorem states that

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)p(\theta)}{f(\boldsymbol{x})}, \tag{2.3.1}$$

where $\pi(\theta|\boldsymbol{x})$ is the posterior density of the parameter, given the observed data, $\boldsymbol{x}$; $f(\boldsymbol{x}|\theta)$ is the likelihood of the data, which is the probability of observing different data, $\boldsymbol{x} = \{x_1, ..., x_n\}$ given that θ is a true parameter; and $p(\theta)$ is the prior density used to evaluate the probability of parameter θ. This prior incorporates our initial belief and uncertainty regarding θ before data are observed. $f(\boldsymbol{x})$ is the normalising constant and by the law of total probability, $f(\boldsymbol{x}) = \int_\theta f(\boldsymbol{x}|\theta)p(\theta)d\theta$. Because this integration can be tedious to calculate and $f(\boldsymbol{x})$ does not involve θ, we can often omit this term in the MCMC techniques described later. Moreover, the posterior $\pi(\theta|\boldsymbol{x})$ only needs to be known up to a constant proportionality for us to make estimation from it. Therefore, the Bayes theorem is often quoted as $\pi(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)p(\theta)$.

We use Bayesian methodology to make inference for parameter θ. Among the useful measures we can obtain from the sampled posterior distribution in making inferences on θ, are the posterior mean $E(\theta|\boldsymbol{x})$, the posterior variance $Var(\theta|\boldsymbol{x})$, and the

Bayesian interval, $\{a(\boldsymbol{x}), b(\boldsymbol{x})\}$ for θ such that $Pr\{a(\boldsymbol{x}) < \theta < b(\boldsymbol{x})|\boldsymbol{x}\} = 0.95$. A 95% Bayesian interval (or credible interval) can be interpreted as 95% chance that the true value of the parameter of interest lays in this interval. There are different types of credible intervals, with two of them being the highest posterior density (HPD) which is the smallest interval enclosing $(1-\alpha)\%$ of the posterior distribution, and the equal-tailed interval which gives 2.5% and 97.5% quantile of the posterior distribution.

2.3.1 Choosing prior distributions

In Bayesian statistics, choosing a parameter prior distribution, $p(\theta)$, is an important task. The prior distribution has to represent our uncertainty about the parameter. With little knowledge of the parameter, either from expert or previous studies, we can use non-informative or flat priors which cover all combinations of parameter values with any support from the likelihood and the requirement is that the prior is relatively flat within this range. For instance, we should assign equal probability to all possible values. Vague prior distributions may cause problems such as large variation in the posterior inference if the sample size is small. With large samples, the effect of changes in prior distributions on results will be minimal.

Jeffreys' prior (Jeffreys, 1961) is an example of non-informative prior which is constructed based on Fisher information, such that

$$I(\theta) = -E_\theta \left[\frac{\partial^2 \log f(\boldsymbol{x}|\theta)}{\partial \theta^2} \right].$$

where $\log f(\boldsymbol{x}|\theta)$ is the usual log-likelihood function of the data $\boldsymbol{x}$.

Therefore, the Jeffrey's prior is defined as

$$p(\theta) \propto [I(\theta)]^{\frac{1}{2}}.$$

If a posterior distribution comes from the same family as the prior distribution, the prior is termed as the conjugate prior for the likelihood. This type of prior approach

was introduced by Schlaifer and Raiffa (1961).

2.4 Markov chain Monte Carlo

As it is often difficult, and in some cases almost impossible, to calculate integrals of a complex posterior distribution, Markov Chain Monte Carlo (MCMC) sampling techniques (Tierney, 1994) are widely used to obtain the posterior distribution iteratively (e.g. Raab et al., 1998). The most well-known MCMC algorithms are Metropolis-Hastings (M-H) and the Gibbs sampler.

2.4.1 Metropolis-Hastings

The Metropolis-Hastings algorithm was developed by Metropolis et al. (1953) and later generalised by Hastings (1970). Let us say we are interested in making inference for parameter $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ can be a vector of parameters, but the posterior distribution is non integrable. The M-H algorithm allows us to generate samples from the target distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ by utilising each parameter's proposal distribution, $q(\boldsymbol{\theta}, .)$.

The steps of the M-H algorithm are summarised as follows.

i. Initialise $\boldsymbol{\theta}^0$ and set $t = 1$.
   For iteration $t = 1, 2, ...$ perform the followings:

ii. (a) Generate proposal value, $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}, .)$ and $U \sim U[0, 1]$.

   (b) Calculate the acceptance probability,

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = \min\{1, \frac{q(\boldsymbol{\theta}^{t-1}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*|\boldsymbol{x})}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{t-1})\pi(\boldsymbol{\theta}^{t-1}|\boldsymbol{x})}\}. \tag{2.4.1}$$

   (c) Accept the proposed value, $\boldsymbol{\theta}^*$ if $U \leq \alpha$. Set $\boldsymbol{\theta}^t = \boldsymbol{\theta}^*$.
   Otherwise, reject the proposed value. Set $\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1}$.

iii. Set $t = t + 1$ and repeat step ii until convergence is achieved.

If the proposal distribution used is a symmetric distribution, i.e. $q(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*) = q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$, the acceptance probability can be simplified to

$$\alpha(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = \min\ \{1, \frac{\pi(\boldsymbol{\theta}^*|\boldsymbol{x})}{\pi(\boldsymbol{\theta}^{t-1}|\boldsymbol{x})}\}. \tag{2.4.2}$$

Note that in calculating the acceptance probability, the normalising constants $f(\boldsymbol{x})$ in (2.3.1) are cancelled out since (2.4.1) is a function involving the ratio of the posterior distribution. Besides updating the parameter one by one, M-H algorithm is also capable of updating a group of parameters at once.

2.4.2 Gibbs sampler

The Gibbs sampler was originally used to analyse the Gibbs distribution on a lattice by Geman and Geman (1984). It was used in statistical physics much earlier, before Gelfand and Smith (1990) and Gelfand et al. (1990) brought MCMC into mainstream. This is a special case of the single-parameter Metropolis-Hastings algorithm. Suppose we have a posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ where $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)'$. Our full conditional distribution, $\pi(\theta_j|\theta_{\setminus j}, \boldsymbol{x})$, $j = 1, ..., p$ is the conditional distribution of θ_j given all other variables where $\theta_{\setminus j} = (\theta_1, ..., \theta_{j-1}, \theta_{j+1}, ...\theta_p)$. With the Gibbs sampler, it is compulsory for the full conditional posterior distribution to be known allowing a random variate to be easily simulated from it. Therefore, we do not need a proposal distribution implying that, since we sample from the posterior distribution to update the j^{th} component of $\boldsymbol{\theta}$, the acceptance probability is always equal to 1. Since the posterior distribution acts as the proposal distribution, we have

$$q(\theta_j|\theta_{\setminus j}, \boldsymbol{x}) = \pi(\theta_1|\theta_{\setminus j}, \boldsymbol{x}). \tag{2.4.3}$$

Thus, substituting (2.4.3) into the acceptance probability in (2.4.1), will make it equal to 1. Therefore, the chain will always move to new values.

The steps of Gibbs sampler algorithm are described as follows.

i. Initialise chain to $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_p^{(0)})$. Set $t = 1$.

ii. Update $t = 1, 2,$ by drawing a sample from each θ component by

$$\theta_1^{(t)} \sim \pi(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)}, x).$$

$$\theta_2^{(t)} \sim \pi(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)}, x).$$

$$\vdots$$

$$\theta_j^{(t)} \sim \pi(\theta_j | \theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)} ..., \theta_p^{(t-1)}, x).$$

$$\vdots$$

$$\theta_p^{(t)} \sim \pi(\theta_p | \theta_1^{(t)}, ..., \theta_{p-1}^{(t-1)}, x).$$

iii. Repeat step 2 until the chain converges.

2.4.3 Convergence diagnostics

To ensure samples come from our distribution of interest, a Markov Chain must run long enough until it converges, i.e. until it reaches the equilibrium or stationary state. However, a common issue with MCMC applications is to determine how many iterations the chain needs before it converges to our target distribution. Although we will never know the exact number of iterations needed, there are several ways to preliminarily assess convergence. Some of them are discussed below.

Examples of quick and simple ways to diagnose convergence are given by calculating the Monte Carlo (MC) error and by monitoring the autocorrelation function of the run (Ntzoufras, 2011). The MC error estimates the difference between the mean of the sampled values and the targeted posterior mean. For instance, a small MC error value implies that the chain has calculated the quantity of interest with precision. Therefore as a rule of thumb, a chain is said to reach its stationary state when the

MC error for each parameter is less than 5% of their standard deviation. As for the autocorrelation, low values indicate fast convergence and vice versa.

We can also monitor the parameters' trace plot, which is a plot of sampled values versus the iterations (Ntzoufras, 2011). If the chain has stabilised around a value without or with little fluctuation, we say that the chain may have reached convergence. Theoretically, we should monitor trace plots for all parameters including parameters that are not of interest. This is because it is possible for some parameters to already mix well while others have bad mixing. Making inference from these sampled values might be inaccurate. However, it is not practical to observe all trace plots for a large parameter set. In this case it is sufficient to randomly check 5 - 10 parameters from a large parameter vector.

A more advanced diagnostic test is to run multiple chains with different starting points (Ntzoufras, 2011). When the lines of these chains cross in the trace plot, the chains are assumed to already have achieved convergence.

In this thesis, we estimate the parameters using MCMC methodology with the aid of WinBUGS (Spiegelhalter et al., 2003). This software has been developed to facilitate the implementation of MCMC methods and now is widely used especially for MCMC simulations from complex statistical models. R software (R Development Core Team, 2009) is also used for other computations and to produce graphical outputs.

2.5 Basic model for individual patients consistency

Based on Zammit et al. (2011), we model the intra-individual consistency using a logistic-type latent variable model. A latent variable is used when a given variable cannot be observed directly and needs to be estimated from a number of related variables. In this case, consistency is non-observable, and also we cannot observe the propensity of symptoms and the intensity of episodes directly. Therefore we need to estimate them through observation of symptoms and episodes of hypoglycaemia.

We define the indicator random variable such that

$$Y_{ijk} = \begin{cases} 1 & \text{if subject } i \text{ reports symptom } j \text{ at episode } k \\ 0 & \text{otherwise.} \end{cases}$$

We assume

$$Y_{ijk} \sim Bernoulli(p_{ijk}) \tag{2.5.1}$$

for individual $i = 1, ..., I$, symptom $j = 1, ..., J$ and episode $k = 1, ..., K_i$, where p_{ijk} is the probability of patient i reporting symptom j at episode k.

Individual i reports symptom j at episode k when $h(\alpha_{ij}, \beta_{ik})$ exceeds a random threshold associated with each patient, where α_{ij} represents the propensity of symptom j for individual i, β_{ik} represents the intensity of episode k for individual i and $h(\cdot)$ is an appropriate functional form. The propensity of a symptom refers to the tendency for a patient to experience that particular symptom, whereas intensity of an episode corresponds to how intense the episode is, with more symptoms experienced in an episode implying higher intensity. The random threshold for each symptom $j = 1, ..., J$ experienced at episode $k = 1, ..., K_i$ by individual $i = 1, ..., I$, denoted by τ_{ijk}, is assumed to follow a log-normal distribution:

$$\tau_{ijk} \sim LN(0, \sigma_i^2). \tag{2.5.2}$$

This analysis is robust to the choice of the threshold distribution. Zammitt et al. (2011) fit the model with thresholds following a Weibull distribution such that, $\tau_{ijk} \sim Weibull(v_i, \lambda_i)$ where v_i and λ_i are the mean and scale parameter respectively and found that the posterior estimates of consistency from the two models are in close agreement.

There is no loss of generality when we set the mean of the log-normal distribution to 0 for all subjects i, because here the mean of the logarithm of the threshold $\mathrm{E}\{\log(\tau_{ijk})\}$ is not of interest and inestimable. Thus, the probability of patient i reporting symp-

tom j at episode k, is given by

$$\begin{aligned} p_{ijk} = & \quad Pr\{\tau_{ijk} \leq h(\alpha_{ij}, \beta_{ik})\} \\ = & \quad \Phi\left\{\frac{\log[h(\alpha_{ij}, \beta_{ik})]}{\sigma_i}\right\}, \end{aligned} \tag{2.5.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable. In earlier work (Zammit et al., 2011) a multiplicative threshold form $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ was assumed, with individual i experiencing symptom j at episode k when $\tau_{ijk} \leq h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. In this thesis we also explore different forms of $h(\cdot)$ as described in Section 3.2.

Parameter σ_i measures the symptom-reporting consistency of a patient. A rescaled consistency parameter, $c_i = \dfrac{100}{(1+\sigma_i^2)}$, is used for easier interpretation where $c_i \in (0, 100]$. A large c_i value indicates high consistency. For large c_i (i.e. small σ_i) the thresholds get highly concentrated around constant reporting of symptoms associated with latent symptom propensity α_{ij} and episode intensity β_{ik} such that $h(\alpha_{ij}, \beta_{ik}) > \tau_i^*$, with τ_i^* approaching a constant value as σ_i tends to zero (or c_i tends to ∞). Therefore, consistent reporting is associated with high concentration of the threshold distribution, corresponding to increased values of the consistency parameter c_i.

Under a Bayesian framework (Berger JO, 1985), we specify appropriate prior distributions for the model parameters α_{ij}, β_{ik} and σ_i. In this analysis, we assume independent priors for the latent variables with

$$\begin{aligned} \alpha_{ij} &\sim Gamma(a_\alpha, b_\alpha), \\ \beta_{ik} &\sim Gamma(a_\beta, b_\beta), \end{aligned} \tag{2.5.4}$$

for individual $i = 1, ..., I$, symptom $j = 1, ..., J$, and episode $k = 1, ..., K_i$. We assign an inverse-gamma prior distribution to the variance parameter, σ_i^2, given as

$$\sigma_i^2 \sim Inv - Gamma(\gamma_\sigma, \delta_\sigma) \tag{2.5.5}$$

for $i = 1, 2, ..., I$.

The likelihood function of this model can be written as

$$f(y_{ijk}|\alpha_{ij}, \beta_{ik}, \sigma_i) \propto \prod_i \prod_j \prod_k \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}^{y_{ijk}} \left[1 - \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}\right]^{1-y_{ijk}}. \tag{2.5.6}$$

The joint posterior distribution of the parameters can be expressed as

$$\pi(\alpha_{ij}, \beta_{ik}, \sigma_i|y_{ijk}) \propto f(y_{ijk}|\alpha_{ij}, \beta_{ik}, \sigma_i)p(\alpha_{ij})p(\beta_{ik})p(\sigma_i), \tag{2.5.7}$$

where $p(\alpha_{ij})$, $p(\beta_{ik})$ and $p(\sigma_i)$ are the prior densities given in (2.5.4) and (2.5.5).

Therefore, from (2.5.7), we can get the full conditional distributions for each parameter:

$$\pi(\alpha_{ij}|\alpha_{\setminus ij}, \beta_{ik}, \sigma_i, y_{ijk}) \propto \prod_k \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}^{y_{ijk}} \left[1 - \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}\right]^{1-y_{ijk}}$$
$$\times \alpha_{ij}^{a_\alpha - 1} \exp(-b_\alpha \alpha_{ij}) \text{ for } i = 1, ..., I, j = 1, ..., J. \tag{2.5.8}$$

$$\pi(\beta_{ik}|\beta_{\setminus ik}, \alpha_{ij}, \sigma_i, y_{ijk}) \propto \prod_j \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}^{y_{ijk}} \left[1 - \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}\right]^{1-y_{ijk}}$$
$$\times \beta_{ik}^{a_\beta - 1} \exp(-b_\beta \beta_{ik}) \text{ for } i = 1, ..., I, k = 1, ..., K_i. \tag{2.5.9}$$

$$\pi(\sigma_i|\alpha_{ij},\beta_{ik},\sigma_{\setminus i},y_{ijk}) \propto \prod_j \prod_k \Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}^{y_{ijk}} \left[1-\Phi\left\{\frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i}\right\}\right]^{1-y_{ijk}}$$

$$\times \sigma_i^{2(-\lambda_\sigma-1)} \exp(-\delta_\sigma/\sigma_i^2) \text{ for } i=1,...,I. \quad (2.5.10)$$

To estimate c_i, we need to estimate the latent variables α_{ij} and β_{ik} first. Estimation of α_{ij} and β_{ik} is informed by the frequency with which a symptom is reported throughout all episodes and the number of symptoms per particular episode. We use a Bayesian approach to estimate the posterior distribution of the unobserved latent factors and the variability of the thresholds. Posterior distributions of the latent variables are obtained using Markov chain Monte Carlo simulation techniques (Arminger and Muthen, 1998). The results, together with relevant discussion and extensions, are given in the following chapters.

Chapter 3

Modelling Individual Patients' Symptom Reporting Consistency

In this chapter, we model individual patients' consistency mainly using Bayesian methodology and Markov chain Monte Carlo techniques. To begin with, based on the core model discussed in Chapter 2, we explore different functional forms, $h(\cdot)$, for the threshold according to which individual i reports symptom j at episode k. Then, we compare the models' fit using Deviance Information Criterion (DIC) before selecting a threshold to use throughout this thesis. Next, we expand the core model by adding variability to symptoms' reporting. Consequently, we perform comparisons on the outcome from each model. Analysis in this chapter will include the data collected from 66 patients including those from seven additional patients that we received later in this study.

3.1 Consistency model for all patients

Referring to the core model in Chapter 2, i.e. (2.5.1) - (2.5.5), Zammit et al. (2011) in earlier work assumed a multiplicative threshold form, $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ in (2.5.3) with individual i experiencing symptom j at episode k when $\tau_i \leq h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Under Bayesian framework, we specify appropriate prior distributions for the model parameters α_{ij}, β_{ik} and σ_i.

We assign the following non-informative prior distributions

$$\alpha_{ij} \sim Gamma(1, 0.1), i = 1, ..., 66 \text{ and } j = 1, ..., 26$$

$$\beta_{ik} \sim Gamma(1, 0.1), i = 1, ..., 66 \text{ and } k = 1, ..., K_i$$

$$\sigma_i^2 \sim Inv - Gamma(1, 0.1), i = 1, ..., 66. \tag{3.1.1}$$

These parameters are estimated by using MCMC methodology implemented in WinBUGS. Posterior estimates for the parameters of interest are computed from 10,000 MCMC iterations after discarding the first 1000 iterations as the burn-in period. Figure 3.1 shows the trace plots displaying the mixing of the chain for the precision parameter, σ_i, for four of the patients: Subjects 1009, 1025, 1028 and 2013. Notice that these chains quickly converge from their initial values (left plots) and are mixing well, reaching their stationary states (right plots). Recall that the consistency parameter c_i is a function of σ_i. Note also that the MC error for each patient (Table 3.1) after 10 000 iterations is less than (or about) 5% of the corresponding parameter standard deviation implying that the number of iterations is appropriate.

The distribution of the posterior mean estimates of the consistency parameter for all 66 patients is shown in Figure 3.2, where estimates of both $\tilde{c}_i$ and $\log(\sigma_i^{-2})$ are presented (left and right respectively). The distribution of the converted consistency, $\tilde{c}_i = \mathrm{E}(\tilde{c}_i|y)$ is presented in a histogram (Figure 3.2 (left)). On the left graph, the red lines refer to the highest and lowest estimated consistency, $\tilde{c}_i$ and the dotted lines refer to their 95% Bayesian intervals (using equal-tailed intervals). Subject 1028, who had Type 1 diabetes for more than 15 years, has the highest estimated consistency (96.69; 95% Bayesian interval (92.95, 98.66)). This patient recorded 45 episodes of hypoglycaemia. The lowest consistency score was estimated for subject 2013 (17.77; 95% Bayesian interval (13.23, 23.02)). This patient had reported the highest number of episodes (138) and had Type 1 diabetes for more than 15 years. The average of $\tilde{c}_i$ is 51.86 with standard deviation 17.19. The main sample quartiles of $\tilde{c}_i$ are q_0=17.77, $q_{0.25}$= 38.46, $q_{0.5}$=51.81, $q_{0.75}$= 63.92 and q_1= 96.67. A list of $\tilde{c}_i$ estimates for all patients is provided in Table 3.1.

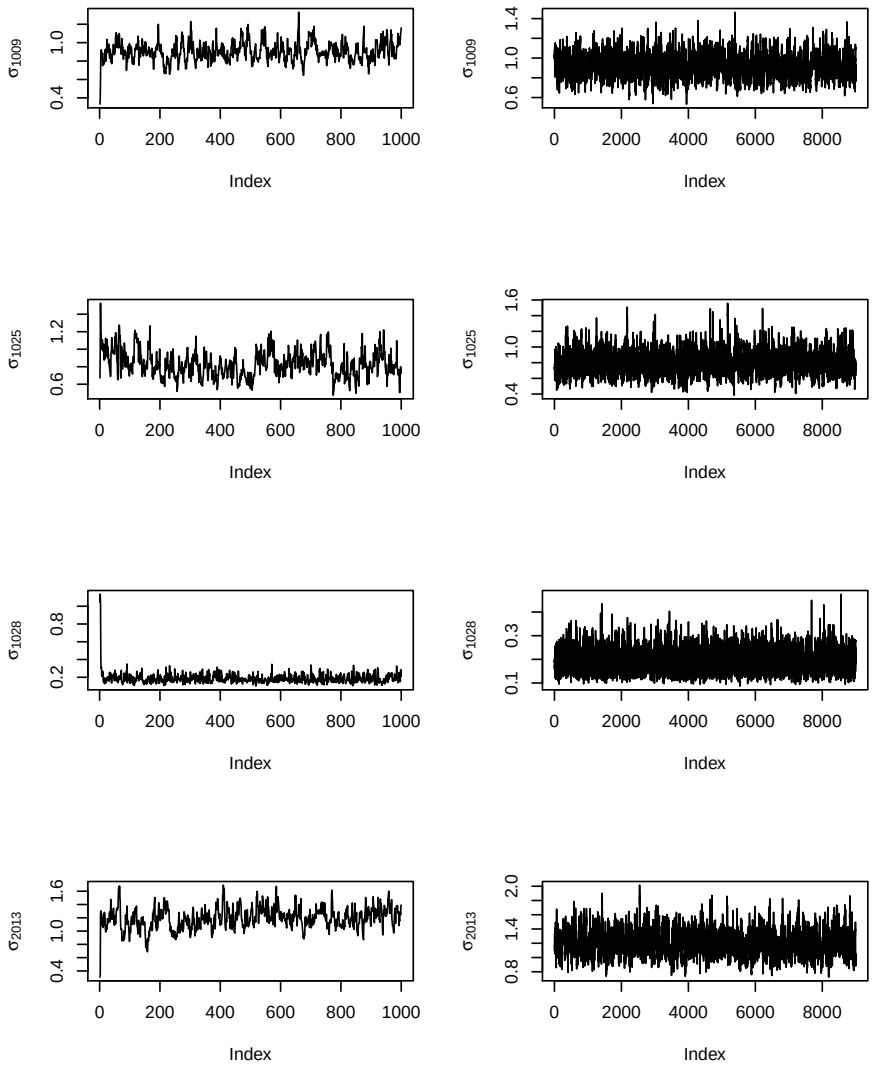


Figure 3.1: Trace plots of parameter σ_i for subjects 1009, 1025, 1028 and 2013 for model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$.

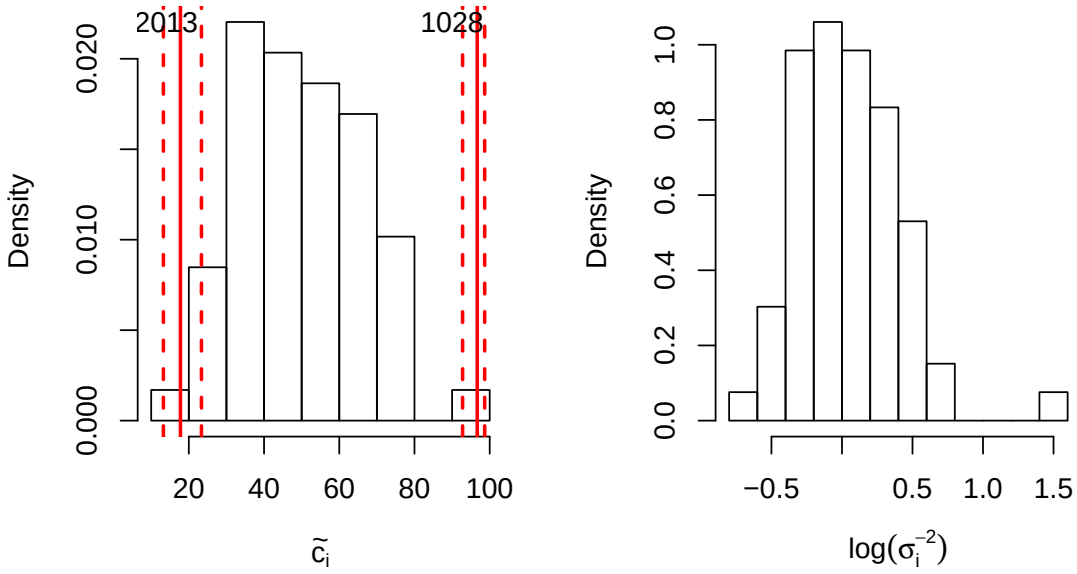


Figure 3.2: Histogram of estimated consistency parameter, $\tilde{c}_i$ and the log estimated precision parameter, σ_i^{-2}.

3.2 Sensitivity to the prior

This section investigates how sensitive are the posterior estimates to variations in the choice of the prior distributions (3.1.1) used in the model. To check this prior sensitivity, we study the posterior estimates of the consistency parameter for all 66 patients using a variety of priors for parameters α_{ij}, β_{ik}, and σ_i^2, which belong to the same family of distributions as those in (3.1.1), but have different prior hyper-parameters, guided mainly by the location and spread of α_{ij}, β_{ik} and σ_i^2. We note here that prior knowledge on likely values of these parameters is not available and therefore our choices here are purely driven by exploratory and comparison purposes. Table 3.2 lists the different prior distributions introduced to evaluate this sensitivity, whereas Table 3.3 gives the combinations of the prior settings used to fit the consistency model.

We obtain estimates of the posterior mean of individual consistency, $\tilde{c}_i$, using each new prior setting and compare them with the results obtained using the original model with prior given in (3.1.1). For our prior sensitivity investigation, we are not inter-

ested in absolute values of consistency, but instead we will examine relative values among patients and focus on the ranking of the individuals' consistency. Therefore, we rank the consistency in descending order and compare the ranking.

Figure 3.3 shows five graphs representing these comparisons. The graphs demonstrate that, in general the consistency ranking of the 66 patients is similar regardless of which prior is used. Correlations in the ranking from each comparison are provided in Table 3.3 and also confirm that the choice of prior does not affect considerably the ranking. We note that the ranking is most affected when prior setting (f) is used, with α_{ij}, $\beta_{ik} \sim$ Gamma(1,1) and $\sigma_i^2 \sim$ Inv-Gamma(1,0.1), resulting in the lowest correlation (0.93). Although this still indicates strong agreement, it is perhaps expected as the Gamma(1,1) prior on α_{ij} and β_{ik} imposes a rather strong assumption of precise prior knowledge (relatively small prior variance).

Subject	No of episodes	$\tilde{c}_i$	SD	95% interval	MC error
1008	95	29.84	4.133	(22.46, 38.26)	0.2239
1009	74	52.21	5.653	(41.40, 63.73)	0.3122
1015	27	36.70	7.275	(23.30, 52.12)	0.3524
1021	43	47.95	6.809	(34.96, 61.66)	0.3322
1025	25	63.28	8.230	(46.96, 78.89)	0.4295
1028	42	96.68	1.522	(92.95, 98.66)	0.0301
1036	24	52.06	8.267	(36.37, 68.82)	0.4073
1039	43	31.03	5.998	(20.63, 43.90)	0.2619
1055	27	74.16	7.270	(58.24, 86.55)	0.3059
1057	11	73.17	8.018	(55.49, 86.73)	0.2468
1086	67	39.95	5.485	(29.43, 50.73)	0.2913
2009	86	30.11	4.207	(22.65, 38.97)	0.2515
2010	89	33.05	4.454	(24.82, 42.26)	0.2413
2012	37	29.59	6.093	(19.06, 42.74)	0.2883
2013	138	17.7	2.507	(13.23, 23.02)	0.1556
2015	32	43.71	6.929	(30.57, 57.41)	0.3219
2021	24	57.32	8.859	(39.92, 73.77)	0.3666
2022	36	44.39	7.482	(30.95, 59.67)	0.3826
2027	22	69.40	7.874	(52.89, 83.71)	0.2953
3001	31	57.54	7.883	(41.76, 72.58)	0.3667
3015	19	48.40	8.852	(31.93, 66.88)	0.3377
3016	22	72.61	7.271	(57.13, 85.41)	0.2926
3022	20	65.86	8.577	(47.66, 80.96)	0.3685
3024	27	50.78	8.147	(35.06, 66.94)	0.4595
3029	69	64.82	6.061	(52.46, 76.06)	0.2324
3043	17	49.06	8.694	(32.27, 66.45)	0.3043
3046	25	69.37	7.780	(52.39, 82.89)	0.3453
3048	31	66.94	7.740	(51.34, 81.66)	0.3162
3050	44	48.96	7.186	(35.57, 63.44)	0.2896
3052	105	49.93	5.298	(39.93, 60.36)	0.2313
3057	51	40.06	6.245	(28.49, 52.95)	0.3188
3065	26	55.14	8.683	(38.10, 71.45)	0.3568
3067	26	42.26	7.906	(27.92, 58.42)	0.3592
4003	23	75.69	6.773	(61.10, 87.31)	0.3045
4008	59	31.17	4.556	(23.21, 40.98)	0.2475
4013	39	74.45	7.192	(59.29, 86.94)	0.3359
4023	134	22.25	3.086	(16.71, 28.57)	0.1643
4028	12	62.02	8.886	(43.93, 78.55)	0.2222
4032	10	84.15	6.954	(68.05, 94.49)	0.1606
4034	55	60.01	6.723	(46.90, 72.76)	0.3282
4043	35	38.67	6.648	(26.61, 52.46)	0.3160
4045	38	48.22	7.483	(33.65, 63.32)	0.3686
4049	22	60.42	8.070	(44.14, 75.82)	0.3563
4061	12	63.55	9.032	(45.25, 80.02)	0.2802
4063	42	34.50	5.926	(23.47, 46.91)	0.2832
4072	39	57.68	7.089	(44.34, 71.78)	0.3651
4076	29	52.73	7.610	(38.08, 67.65)	0.4447
5004	26	59.22	8.468	(42.49, 75.38)	0.3641
5009	36	26.18	5.081	(17.58, 37.52)	0.2674
5023	20	72.88	7.750	(56.47, 86.41)	0.3645
5026	24	74.44	7.341	(58.49, 87.07)	0.3055
5029	45	62.37	6.780	(49.03, 75.35)	0.2959
5044	70	37.62	4.739	(28.70, 47.22)	0.2302
5045	68	47.19	5.708	(36.29, 58.70)	0.2595
5088	87	23.98	3.527	(17.70, 31.43)	0.1964
6002	34	40.89	6.630	(29.09, 54.72)	0.3843
6010	19	64.27	8.196	(47.84, 79.65)	0.4192
6018	39	29.63	5.642	(20.03, 42.03)	0.3078
6019	64	38.30	5.716	(27.97, 50.08)	0.2313
6023	59	32.35	4.774	(23.68, 42.19)	0.2312
6038	79	28.63	4.051	(21.36, 37.29)	0.2227
6062	13	59.60	9.485	(40.82, 77.35)	0.3914
6064	17	85.07	5.532	(72.48, 93.77)	0.3147
6065	39	57.41	6.479	(44.58, 69.58)	0.2890
6056	28	49.88	7.411	(35.97, 64.51)	0.1541
6058	20	75.02	6.943	(60.59, 87.41)	0.3049

Table 3.1: Posterior estimates of consistency (Mean, standard deviation, 95% Bayesian interval and Monte-Carlo error) for all subjects using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$.

α_{ij}, β_{ik} priors	$E(\alpha)$	$Var(\alpha)$	$E(\beta)$	$Var(\beta)$
$\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.1)$	10	100	10	100
$\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.01)$	100	10^4	100	10^4
$\alpha_{ij}, \beta_{ik} \sim Ga(1, 1)$	1	1	1	1
σ_i^2 priors	$E(\sigma^2) = \frac{\beta}{\alpha-1}$	$\text{mode}(\sigma^2) = \frac{\beta}{\alpha+1}$	$Var(\sigma^2) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$	
$\sigma_i^2 \sim Inv - Ga(1, 0.1)$	∞	0.05	∞	
$\sigma_i^2 \sim Inv - Ga(2.1, 10)$	9.1	3.23	826.4	
$\sigma_i^2 \sim Inv - Ga(2.1, 1)$	0.9	0.32	8.26	
$\sigma_i^2 \sim Inv - Ga(1.1, 10)$	100	4.76	∞	

Table 3.2: Alternative priors explored.

Prior	Correlation
(a) $\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.1)$ and $\sigma_i^2 \sim Inv - Ga(1, 0.1)$	-
(b) $\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.1)$ and $\sigma_i^2 \sim Inv - Ga(2.1, 10)$	0.96
(c) $\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.1)$ and $\sigma_i^2 \sim Inv - Ga(2.1, 1)$	0.99
(d) $\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.1)$ and $\sigma_i^2 \sim Inv - Ga(1.1, 10)$	0.96
(e) $\alpha_{ij}, \beta_{ik} \sim Ga(1, 0.01)$ and $\sigma_i^2 \sim Inv - Ga(1, 0.1)$	0.98
(f) $\alpha_{ij}, \beta_{ik} \sim Ga(1, 1)$ and $\sigma_i^2 \sim Inv - Ga(1, 0.1)$	0.93

Table 3.3: New prior settings compared to original prior (a) and the correlations of the ranking of consistency estimates, $\tilde{c}_i$ between them.

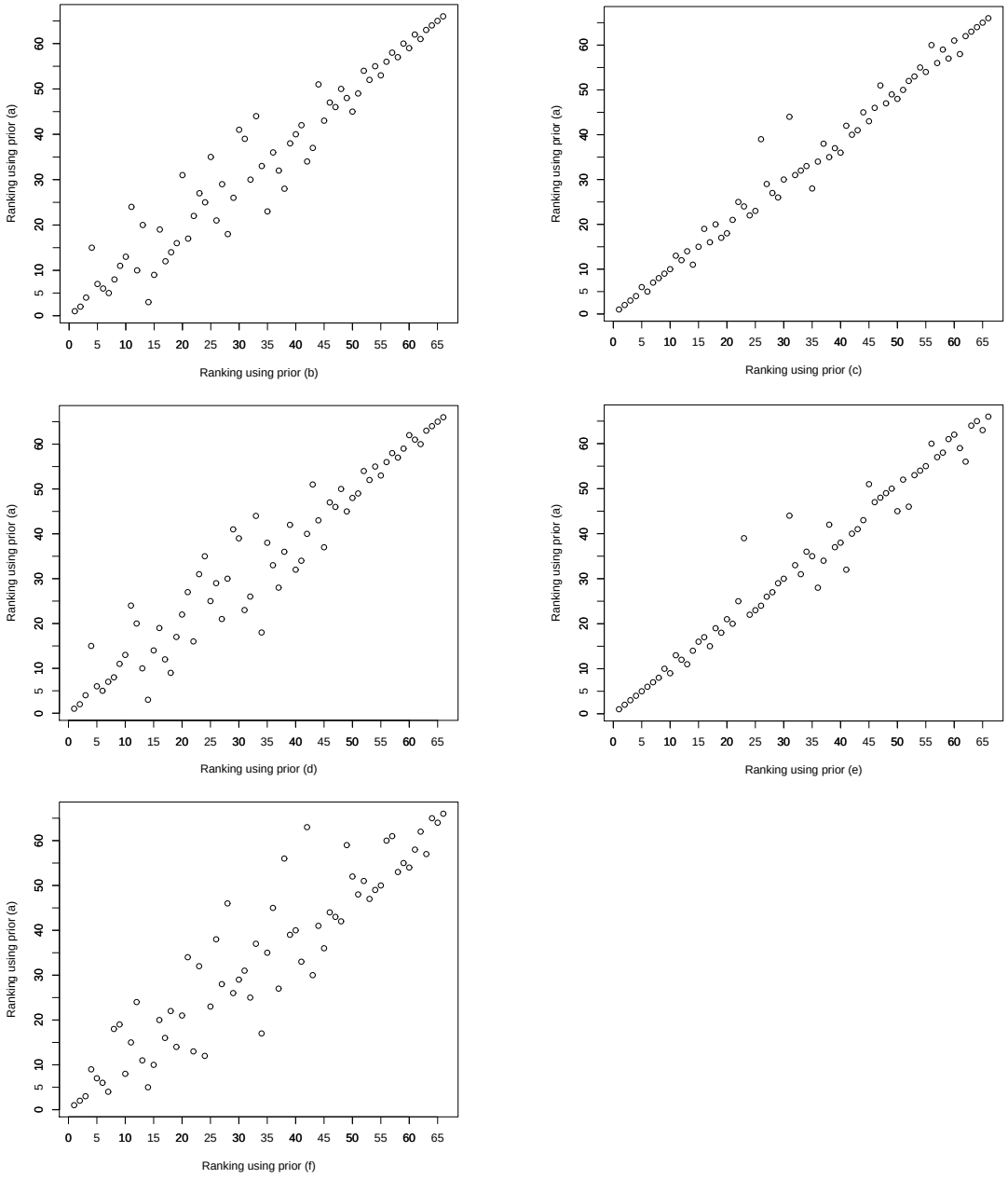


Figure 3.3: Ranking of consistency estimates, $\tilde{c}_i$, of 66 patients when using different prior settings - see Table 3.3.

3.3 Exploration of different thresholds

In this section, we also explore two other different thresholds to fit to this model. These are:

i. $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$, and

ii. $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$.

For each new threshold, the probability of patient i reporting symptom j at episode k is obtained by substituting the thresholds into (2.5.3). Figures 3.4-3.5 show the trace plots of parameter σ_i for subjects 1009, 1025, 1028 and 2013 when using thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$ and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ respectively, for 10,000 iterations. The first 1000 iterations (left figures) are the burn-in period. However, we can see that the chains quickly converge to some specific values within the first 200 iterations. The estimates of consistency for each patient for these two models are summarised in Tables 3.4 and 3.5. The distributions of the estimated consistency $\widetilde{c}_i$ and the logarithm of estimated precision parameter σ_i^{-2} for both models are given in Figure 3.6.

Using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$, the majority of the subjects have high consistency with estimated mean of converted consistency parameter $\tilde{c}_i = 86.25$ and standard deviation of 5.4369. Posterior estimates of $\tilde{c}_i$ and its credible intervals for each subject are displayed in Table 3.4. Subject 1028 has the highest consistency score, $\tilde{c}_i = 98.20$ with 95% Bayesian interval (96.76, 99.08). Although subject 2013 still gives the lowest estimated $\tilde{c}_i$, the value now gets drastically higher compared to the values estimated from our previous model in Section 3.1 (73.26; 95% Bayesian interval (67.93, 78.44)). Subject 3029 had the second highest consistency score (94.97; 95% Bayesian interval (92.57, 96.76)). However, this patient was previously ranked at 13^{th} largest consistency estimate. The patient had diabetes Type 1 for less than five years and reported 69 episodes of hypoglycaemia.

Dropping the interaction term $\alpha_{ij}\beta_{ik}$, we now have $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$. Again, the majority of the subjects have high consistency with estimated mean of converted consistency parameter $\overline{\widetilde{c}_i} = 89.47$ and standard deviation of 4.1310. The summary

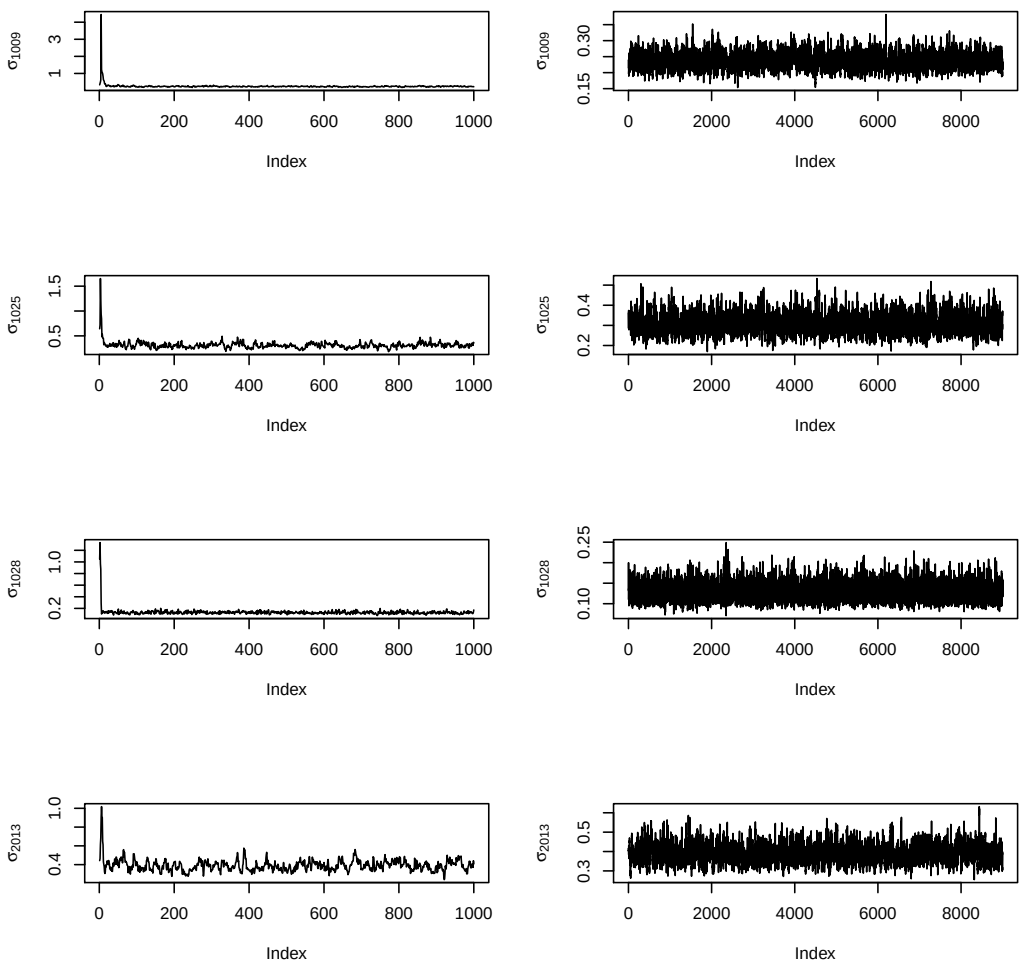


Figure 3.4: Trace plots of parameter σ_i for subjects 1009, 1025, 1028 and 2013 using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$.

statistics of the posterior consistency estimates, $\tilde{c}_i$, for each patient are given in Table 3.5. Similar to the result from the previous thresholds, subject 1028 still has the highest consistency, with $\tilde{c}_i = 98.36$ and 95% Bayesian interval (97.08, 99.13). The subject with the lowest consistency is 2013 with $\tilde{c}_i = 78.69$ and 95% Bayesian interval (73.77, 82.96).

Estimating individual patients' consistency, $\widetilde{c}_i$, using the model with threshold $h(\alpha, \beta) = \alpha_{ij}\beta_{ik}$ clearly yielded a much wider range of $\widetilde{c}_i$ compared to the patients' consistency estimated from the other two thresholds (Figure 3.7). Although the scale and spread of $\widetilde{c}_i$ has changed, generally consistency ranking of the patients is mostly similar except for a few patients who have quite big differences (Figures 3.8-3.9).

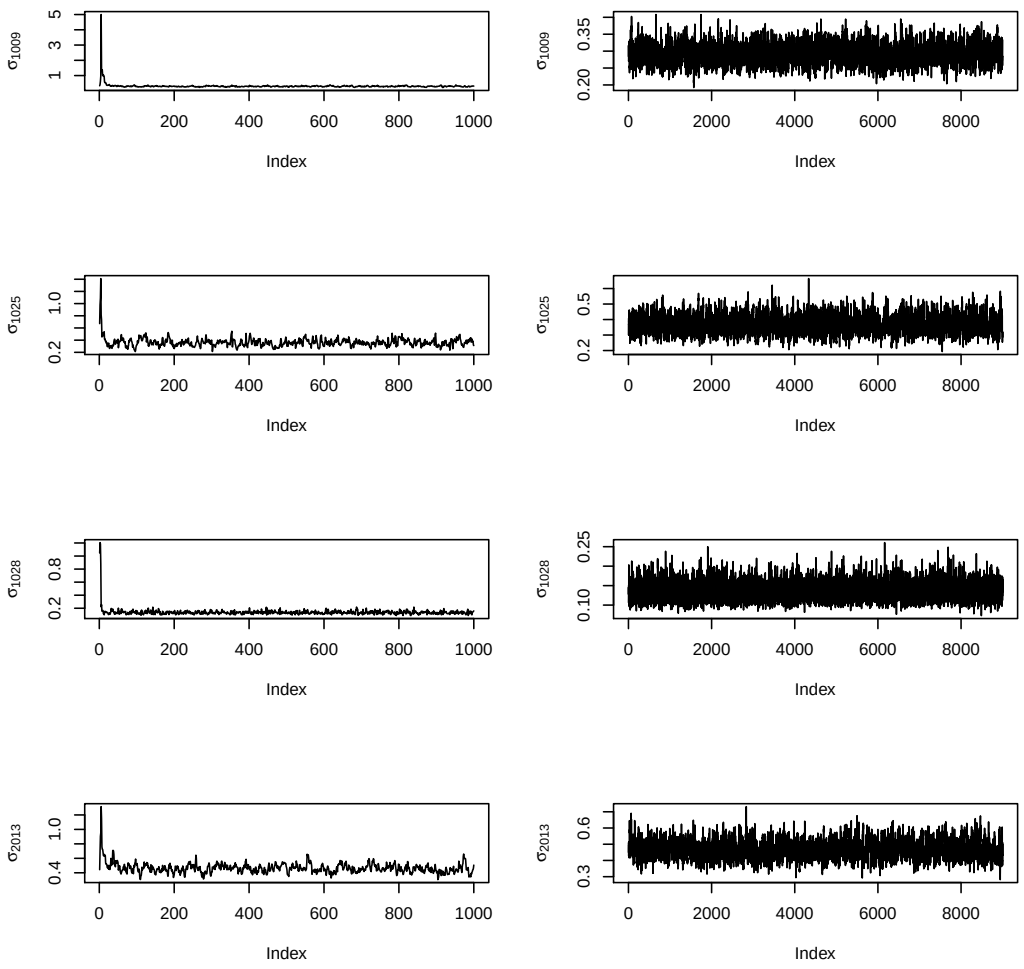
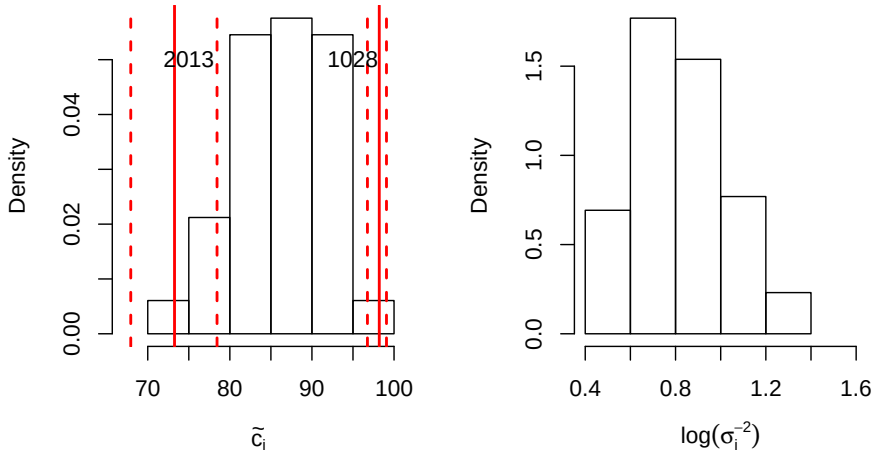


Figure 3.5: Trace plots of parameter σ_i for subjects 1009, 1025, 1028 and 2013 using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$.

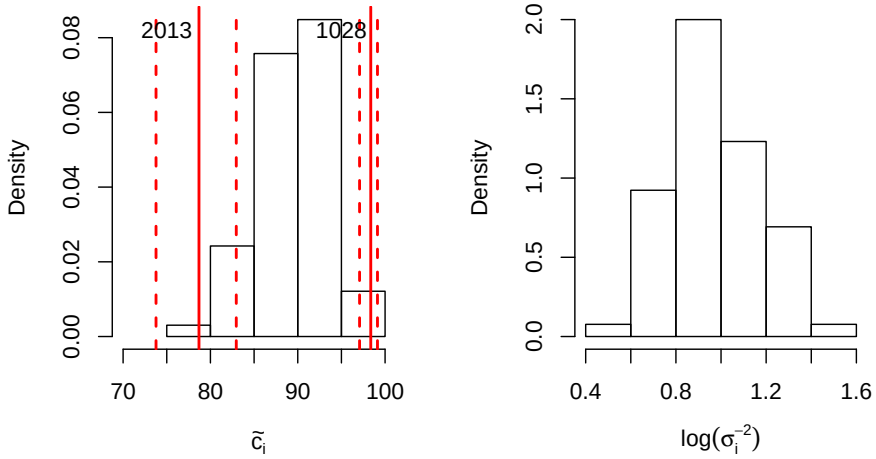
Note that these patients have a small number of episodes. We are not interested in the absolute values of this estimate as we cannot define a cut-off to determine which subject is consistent and which is not. Therefore, what is more important here is how consistent a particular patient is, compared to other patients.

3.3.1 Model selection: DIC criterion

To compare models with different thresholds, we first use the Deviance Information Criterion. DIC was proposed as a Bayesian measure of model fit by Spiegelhalter et al. (2002), following the principle that it is a normal way to compare models using a criterion based on the trade-off between the fit of the data to the model and the



(a) Model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$.



(b) Model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$.

Figure 3.6: Histograms of estimated consistency parameter, $\tilde{c}_i$, and the log of estimated precision parameter, σ_i^{-2}, for model with different thresholds.

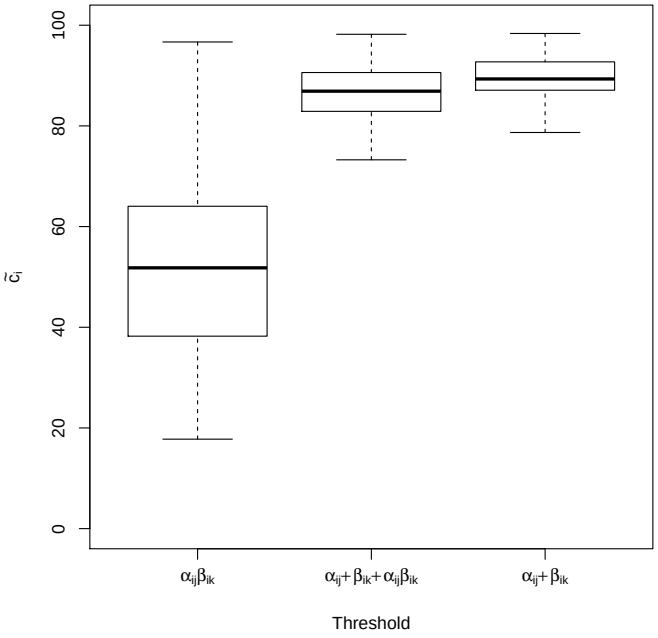


Figure 3.7: Boxplots of estimated consistency parameter, $\tilde{c}_i$ in model with different thresholds.

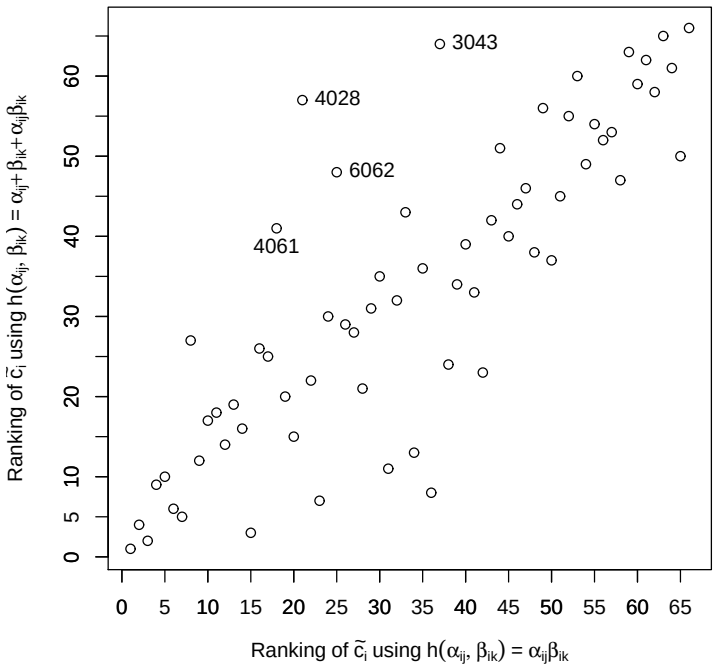


Figure 3.8: Ranking of consistency estimates, $\tilde{c}_i$, of 66 patients for model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$, versus model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$.

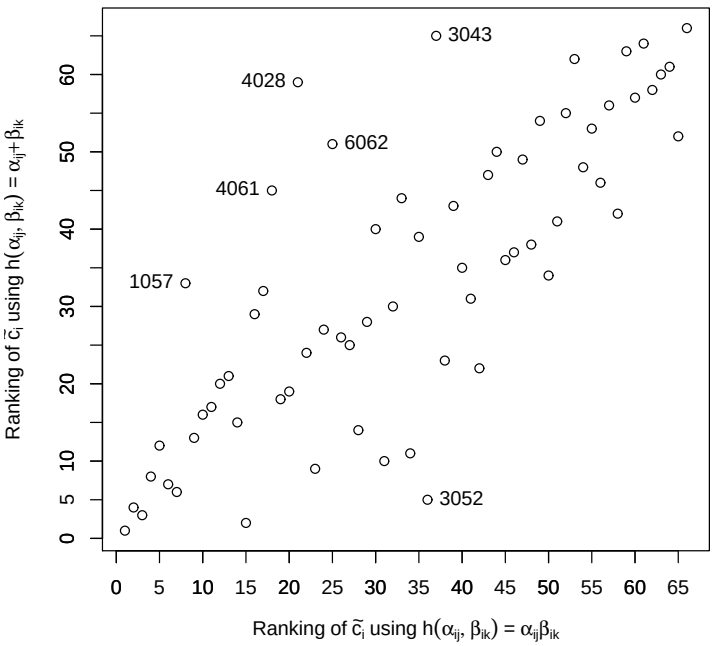


Figure 3.9: Ranking of consistency estimates, $\tilde{c}_i$, of 66 patients for model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$, versus model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$.

corresponding complexity of the model. Thus, DIC is given by

$$\text{DIC} = \text{“goodness-of-fit”} + \text{complexity}.$$

Goodness-of-fit is measured by deviance, $D(\theta)$, which is a measure of overall fit for a likelihood, $L(\theta; y)$. Following Gelman et al. (2004) the deviance is defined as

$$D(\theta) = -2logL(\theta; y).$$

Complexity is measured by the effective numbers of the parameters and is defined as

$$d_e = \widetilde{D(\theta)} - D(\widetilde{\theta}) \tag{3.3.1}$$

where $\widetilde{D(\theta)}$ is the posterior mean of $D(\theta)$ with respect to the posterior distribution of θ. $D(\widetilde{\theta})$ is $D(\theta)$ evaluated at the posterior mean of θ. Therefore DIC is given by

$$DIC = 2\widetilde{D(\theta)} - D(\widetilde{\theta})$$

$$= D(\widetilde{\theta}) + 2d_e.$$

Larger deviance indicates poor fit. Hence, a model with smaller DIC value provides better fit of the data. In general, if a model's DIC value exceeds another model's DIC value by 5 or more, we can rule out that model, but if the difference of DIC reported by the models is less than 5 and both models yield very different inferences, it is not reliable to make conclusion based only on the DIC (Spiegelhalter et al., 2002).

We start our comparison by examining the DIC values of the three thresholds used in the model on eight subjects individually. The following subjects were randomly chosen from all 66 subjects: 1009, 1036, 3016, 3046, 3052, 3057, 3065 and 5009. Smaller DIC values indicate a better model. From the results summarised in Table 3.6, it is worth noticing that subjects with small number of episodes favour threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}, \beta_{ik}$ whereas subjects who experienced larger number of episodes favour a model with $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$. The differences between the DIC values from different thresholds for Subject 3016 are less than five, suggesting that all three models fit the data reasonably well and it is not sufficient to distinguish the better model based solely on the DIC values. We further our study by comparing DIC values for the models with all subjects and summarise the results in the same table. The model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ produced the lowest DIC value suggesting that this gives the best fit to the data. Therefore, all further analysis in this thesis will be using this threshold. This is also supported by an analysis based on a different comparison approach, presented in Section 3.5.

Table 3.6 elaborates more on the DIC results of these models by showing the values of the deviance and the effective number of parameters together with the DIC values corresponding to each model. Note that the number of parameters of the model with all patients is also given in Table 3.6. The results in Table 3.6 show that the value of effective number of parameters differs considerably between models for cer-

tain patients, which may not be expected here. For example, the effective number of parameters for patients 3046 and 3065 seems uncharacteristically low under the model with threshold $\alpha_{ij}\beta_{ik}$. This highlights the problems with DIC methodology, as discussed in Spiegelhalter et al. (2002) and Spiegelhalter et al. (2014). Therefore, in order to confirm these model selection findings, we also present in the next section results based on maximum likelihood estimation.

3.3.2 Model selection: Maximum likelihood estimation

We fit the models with the three different thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$; $\alpha_{ij}+\beta_{ik}$; $\alpha_{ij}+\beta_{ik}+\alpha_{ij}\beta_{ik}$ using maximum likelihood estimation. For convenience we consider the log-likelihood function, and using the likelihood in (2.5.6), our log-likelihood function for the model with all patients when using threshold $h(\alpha_{ij}), \beta_{ik} = \alpha_{ij}\beta_{ik}$ becomes

$$\log f(y_{ijk}|\alpha_{ij}, \beta_{ik}, \sigma_i) \propto \sum_i \sum_j \sum_k (y_{ijk}) \log \Phi \left\{ \frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i} \right\}$$

$$+(1 - y_{ijk}) \log \left[1 - \Phi \left\{ \frac{\log(\alpha_{ij}\beta_{ik})}{\sigma_i} \right\} \right].$$

For comparison purposes we also present in this section log-likelihood results for the same individual patients presented in Table 3.6 when the DIC method was used.

We maximise the log-likelihood function above in R using the 'nlminb' command which minimises the negative log-likelihood function. Here, in this classical analysis, the form of the likelihood gives identifiability issues, with parameters α_{ij}, β_{ik} and σ_i not all being identifiable in the estimation. So we need to use certain constraints, and we choose to apply the following constraints to the parameters: $\alpha_{i1} = 1$ and $\sigma_i = 1$. We note here that the constraint on σ_i means that the present model can not be used in the classical approach for the purposes of this work, which is to estimate the patients' consistency based on σ_i. However, under the Bayesian approach used in this thesis, identifiability is not an issue as the prior distributions serve as constraints on the parameters by 'pulling' them towards certain values and providing proper poste-

rior distributions for all parameters (Xie and Carlin, 2006). This is the case when the simple independent priors in (3.1.1) are used in the model, but also when common priors are used for groups of symptoms later in the thesis (in Section 3.4).

Table 3.7 summarises the results of the log-likelihood function, evaluated at maximum likelihood estimates of the model parameters, for the three threshold models. Larger values of the log-likelihood function indicate a better model. From the table, we can see that data from the eight patients favour the model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$, which is also favoured when the analysis with all patients is considered. Note that our results in the previous section, when we compare models using DIC values, indicate that the lowest DIC values for patients 1009, 3057, and 3052 are when using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$. However, the overall results are consistent in identifying $\alpha_{ij}\beta_{ik}$ as the most suitable model.

We then look further at the results from fitting the models using the Bayesian method and maximum likelihood. Figure 3.10 compares the estimates of parameter $\alpha_{5009,j}$ for patient 5009 for model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Although the absolute values of the estimates are different, their rankings are the same except for those symptoms with propensity estimates very close to zero. This observation is also true for parameter β_{ik} for patient 5009 (Figure 3.11). The ranking of episodes with higher intensity are the same regardless of the method used. Figures 3.12 and 3.13 compare the estimates of parameters (α_{ij} and β_{ik} respectively for patient 1036 when using maximum likelihood (left) and the posterior estimates from the Bayesian method (right)). We also compare the ranking of the α_{ij} estimates for patient 1009 and 3057 (Figures 3.14 - 3.15). Again, the ranking of the estimated values with maximum likelihood and the posterior estimates under Bayesian estimation are the same for non-negligible estimates. Ranking for β_{ik} estimates also have the same trend (not shown here).

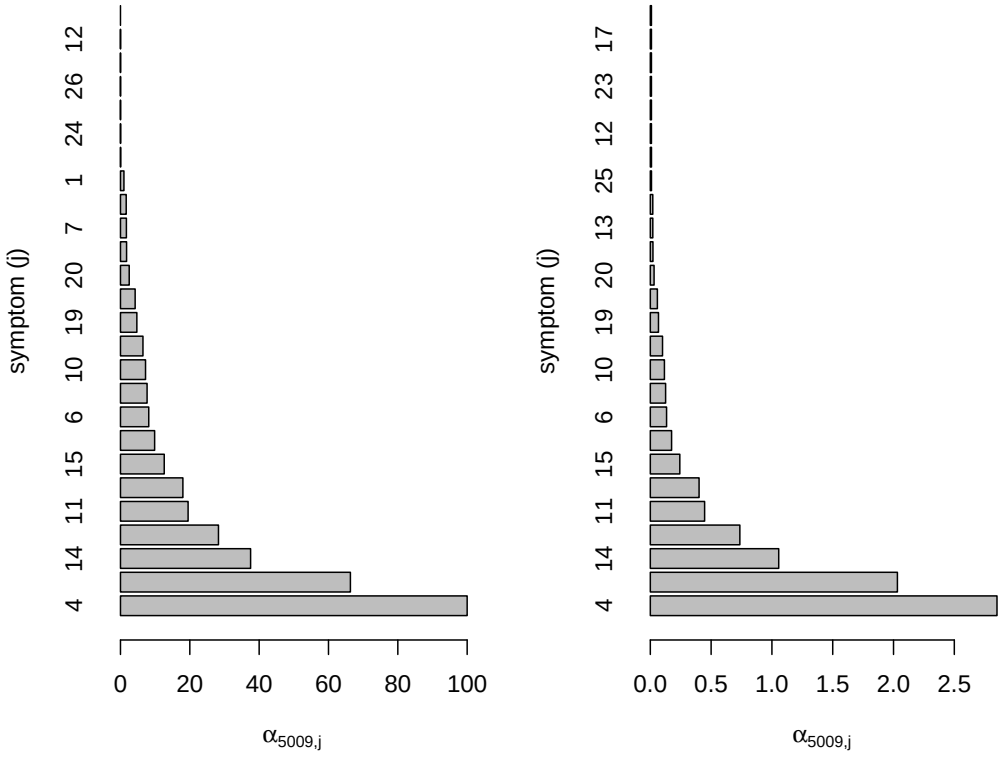


Figure 3.10: Estimates of $\alpha_{5009,j}$ obtained with the maximum likelihood approach (left) and posterior estimates of $\alpha_{5009,j}$ obtained from Bayesian approach (right).

Group, Subject	No of episodes	$\tilde{c}_i$	SD	95% interval
1008	95	83.22	2.289	(78.40, 87.30)
1009	74	91.87	1.477	(88.68, 94.39)
1015	27	81.33	4.031	(72.50, 88.29)
1021	43	86.98	2.709	(81.07, 91.60)
1025	25	88.59	3.150	(81.46, 93.79)
1028	42	98.20	0.605	(96.76, 99.08)
1036	24	91.39	2.314	(86.14, 95.17)
1039	43	77.85	3.917	(68.67, 83.89)
1055	27	92.23	2.317	(86.96, 95.92)
1057	11	88.37	4.884	(76.51, 95.28)
1086	67	85.02	2.461	(79.90, 89.43)
2009	86	81.51	2.501	(76.37, 86.20)
2010	89	82.92	2.273	(78.15, 87.06)
2012	37	77.04	4.151	(68.50, 84.51)
2013	138	73.26	2.652	(67.93, 78.44)
2015	32	84.55	3.354	(77.45, 90.50)
2021	24	88.17	3.300	(80.60, 93.55)
2022	36	82.47	3.544	(74.94, 88.79)
2027	22	90.46	2.886	(83.92, 95.06)
3001	31	87.55	3.063	(80.97, 92.70)
3015	19	86.81	3.968	(78.04, 93.40)
3016	22	91.75	2.632	(85.65, 95.96)
3022	20	90.62	2.870	(84.13, 95.19)
3024	27	87.16	3.242	(79.95, 92.48)
3029	69	94.97	1.077	(92.57, 96.76)
3043	17	75.32	6.363	(61.20, 86.30)
3046	25	90.81	2.562	(85.12, 95.10)
3048	31	90.20	2.659	(84.23, 94.57)
3050	44	88.60	2.414	(83.44, 92.80)
3052	105	92.43	1.179	(89.88, 94.54)
3057	51	83.38	2.870	(77.40, 88.49)
3065	26	85.70	3.542	(78.09, 91.82)
3067	26	84.96	3.687	(76.62, 90.98)
4003	23	92.43	2.149	(87.34, 95.84)
4008	59	81.86	2.777	(75.90, 86.84)
4013	39	85.38	3.084	(78.78, 90.77)
4023	134	82.89	1.827	(79.05, 86.22)
4028	12	81.15	6.295	(66.83, 91.37)
4032	10	95.20	2.077	(90.00, 98.01)
4034	55	92.53	1.611	(88.98, 95.24)
4043	35	81.27	3.693	(73.21, 87.65)
4045	38	85.01	3.187	(78.11, 90.61)
4049	22	89.02	3.107	(82.10, 94.18)
4061	12	84.75	6.170	(69.54, 93.52)
4063	42	78.00	3.860	(70.10, 85.10)
4072	39	88.20	2.671	(82.22, 92.69)
4076	29	84.45	3.663	(76.69, 90.77)
5004	26	87.73	3.319	(80.27, 93.23)
5009	36	75.00	4.624	(65.19, 83.11)
5023	20	90.60	2.861	(84.08, 95.21)
5026	24	92.83	2.222	(87.64, 96.34)
5029	45	90.74	2.116	(86.10, 94.38)
5044	70	83.89	2.514	(78.48, 88.37)
5045	68	88.83	1.898	(84.79, 92.18)
5088	87	77.78	2.781	(72.00, 82.96)
6002	34	84.42	3.454	(76.87, 90.26)
6010	19	88.51	3.409	(80.86, 94.18)
6018	39	79.33	3.961	(70.64, 86.07)
6019	64	85.20	2.293	(80.36, 89.26)
6023	59	80.45	3.014	(74.94, 86.73)
6038	79	79.70	2.947	(73.52, 85.03)
6056	28	90.11	2.831	(83.68, 94.75)
6058	20	92.70	2.269	(87.43, 96.24)
6062	13	83.09	5.376	(71.50, 91.74)
6064	17	94.26	2.212	(88.85, 97.46)
6065	39	89.68	2.257	(84.74, 93.46)

Table 3.4: Posterior estimates of consistency (Mean, standard deviation, and 95% Bayesian Interval) for all subjects using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$.

Group, Subject	No of episodes	$\tilde{c}_i$	SD	95% interval
1008	95	87.83	1.780	(84.25,90.87)
1009	74	93.94	1.194	(91.28,95.93)
1015	27	85.98	3.473	(78.08,91.63)
1021	43	90.81	2.057	(86.10,93.98)
1025	25	91.51	2.409	(86.44,95.49)
1028	42	98.43	0.544	(97.18,99.25)
1036	24	94.10	1.796	(89.51,96.74)
1039	43	82.91	3.473	(74.77,88.72)
1055	27	94.19	1.715	(90.01,96.83)
1057	11	90.94	4.238	(80.34,96.43)
1086	67	89.08	1.996	(84.59,92.98)
2009	86	86.52	2.059	(82.65,90.70)
2010	89	87.12	1.778	(83.33,90.16)
2012	37	81.94	3.580	(74.45,88.10)
2013	138	78.94	2.538	(73.98,84.07)
2015	32	87.34	3.005	(81.15,91.90)
2021	24	91.13	2.532	(85.19,95.29)
2022	36	87.45	2.705	(82.03,92.42)
2027	22	93.12	2.094	(88.52,96.53)
3001	31	91.03	2.186	(85.98,94.85)
3015	19	87.45	3.616	(78.91,93.12)
3016	22	93.55	2.177	(88.43,96.83)
3022	20	93.16	2.093	(88.40,96.32)
3024	27	91.07	2.455	(85.44,94.81)
3029	69	96.61	0.715	(95.18,97.91)
3043	17	81.35	5.524	(68.90,90.40)
3046	25	93.38	1.548	(90.25,96.21)
3048	31	93.02	1.970	(88.62,96.33)
3050	44	92.22	1.820	(88.18,95.22)
3052	105	95.04	0.858	(93.15,96.61)
3057	51	88.14	2.268	(83.44,91.94)
3065	26	89.84	2.562	(84.38,93.83)
3067	26	89.13	3.289	(81.60,93.89)
4003	23	94.50	1.656	(90.77,97.05)
4008	59	87.50	2.345	(82.66,91.54)
4013	39	94.73	1.896	(90.59,97.54)
4023	134	87.80	1.480	(84.77,90.16)
4028	12	84.93	5.471	(72.16,93.10)
4032	10	95.88	1.721	(91.73,98.24)
4034	55	95.07	1.201	(92.29,97.01)
4043	35	85.75	3.189	(78.59,91.21)
4045	38	89.11	2.656	(82.85,93.43)
4049	22	91.91	2.574	(86.46,96.00)
4061	12	88.13	5.292	(75.22,95.29)
4063	42	83.48	3.061	(77.30,88.57)
4072	39	91.60	1.940	(87.42,94.80)
4076	29	87.52	3.045	(80.97,93.11)
5004	26	90.39	2.895	(83.86,94.91)
5009	36	82.27	3.725	(73.86,88.79)
5023	20	92.48	2.369	(87.12,96.18)
5026	24	94.53	1.846	(90.23,97.26)
5029	45	93.54	1.487	(90.01,96.14)
5044	70	87.75	2.201	(82.83,91.67)
5045	68	92.44	1.491	(88.81,94.80)
5088	87	82.56	2.478	(77.05,86.55)
6002	34	88.05	3.104	(80.20,93.04)
6010	19	91.79	2.640	(86.34,96.20)
6018	39	84.89	3.489	(77.12,90.66)
6019	64	89.41	1.699	(85.97,92.90)
6023	59	86.22	2.455	(81.01,90.81)
6038	79	84.98	2.640	(78.78,88.97)
6056	28	88.65	2.558	(83.17,93.02)
6058	20	94.71	1.668	(90.86,97.48)
6062	13	86.49	4.655	(75.31,93.90)
6064	17	95.55	1.689	(91.47,97.94)
6065	39	93.07	1.704	(89.38,95.61)

Table 3.5: Posterior estimates of consistency (Mean, standard deviation, and 95% Bayesian interval) for all subjects using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta ik$.

Patient	No. of episodes	No. of parameters	$D(\widetilde{\theta})$	d_e	DIC
$\alpha_{ij}\beta_{ik}$					
5009	36	63	592.3	28.3	620.6***
1036	24	51	308.8	17.3	326.1***
3016	22	49	154.4	14.4	168.8***
3046	25	52	184.4	7.7	192.0***
3065	26	53	301.6	2.7	304.4***
1009	74	101	313.5	58.6	372.1
3057	51	77	571.0	50.8	621.8
3052	105	131	354.8	67.7	422.5
all	2791	4573	24360.3	2948.1	30256.5***
$\alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$					
5009	36	63	586.4	44.8	631.2
1036	24	51	313.2	38.4	351.5
3016	22	49	148.9	24.8	173.7
3046	25	52	174.4	31.2	205.6
3065	26	53	296.8	36.8	333.5
1009	74	101	298.1	61.0	359.1
3057	51	77	562.0	55.1	617.1
3052	105	131	346.7	78.8	425.5
all	2791	4573	24411.3	3087.1	30585.6
$\alpha_{ij} + \beta_{ik}$					
5009	36	63	591.4	44.3	635.6
1036	24	51	315.7	37.9	353.6
3016	22	49	146.4	25.0	171.4
3046	25	52	172.1	29.4	201.5
3065	26	53	295.5	35.7	331.2
1009	74	101	288.5	57.0	345.5***
3057	51	77	557.7	52.7	610.4***
3052	105	131	335.2	70.9	406.1***
all	2791	4573	25012.9	2725.4	30463.6

Table 3.6: Value of the deviance ($D(\widetilde{\theta})$), effective number of parameter (d_e), and DIC values (*** marks the lowest value) for models with different thresholds, $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$; $\alpha_{ij} + \beta_{ik}$; $\alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$.

Patient	No. of episodes	No. of parameters	Threshold, $h(\alpha_{ij}, \beta_{ik})$ $\alpha_{ij}\beta_{ik}$	$\alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$	$\alpha_{ij} + \beta_{ik}$
5009	36	61	-265.7626^{***}	-299.6110	-298.6920
1036	24	49	-130.4921^{***}	-169.5425	-174.3250
3016	22	47	-57.9646^{***}	-78.7651	-78.5283
3046	25	50	-72.0928^{***}	-92.9805	-93.0193
3065	26	51	-128.7100^{***}	-152.4640	-152.3290
1009	74	99	-119.8340^{***}	-512.7560	-171.5770
3057	51	75	-251.4230^{***}	-254.5510	-254.9340
3052	105	129	-131.4360^{***}	-402.5420	-162.8760
All	2791	4571	-11986.18307^{***}	-14834.30723	-13937.02632

Table 3.7: Maximum likelihood estimates (*** marks the highest value) for individual patients and for all patients.

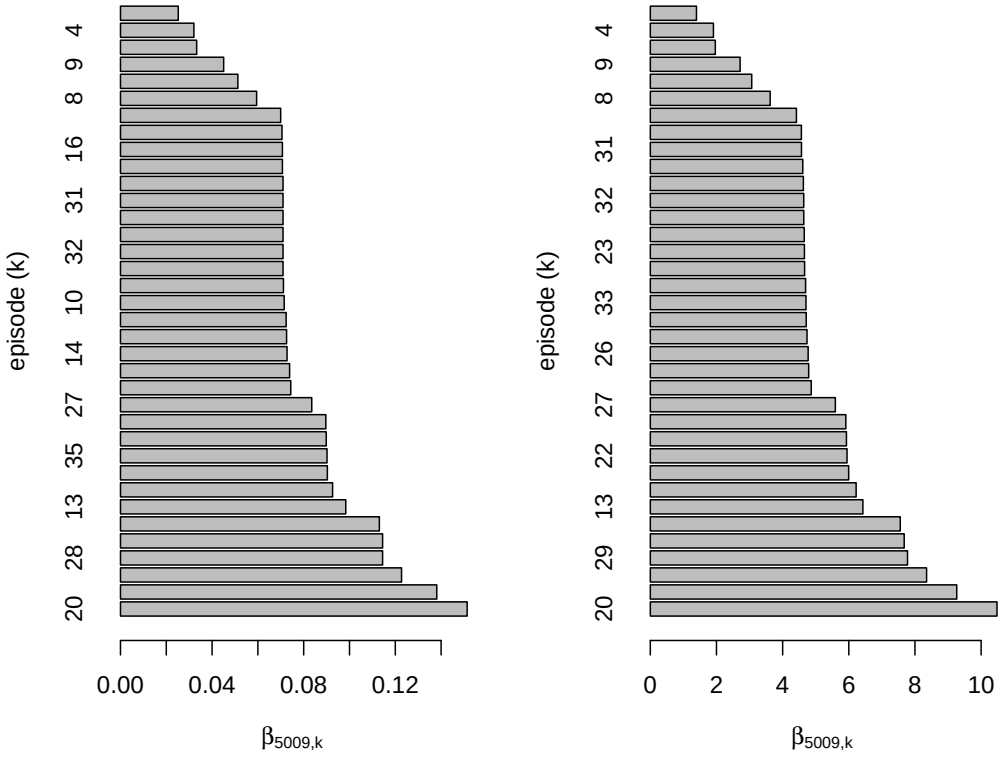


Figure 3.11: Estimates of $\beta_{5009,k}$ obtained with the maximum likelihood approach (left) and posterior estimates of $\beta_{5009,k}$ obtained from Bayesian approach (right).

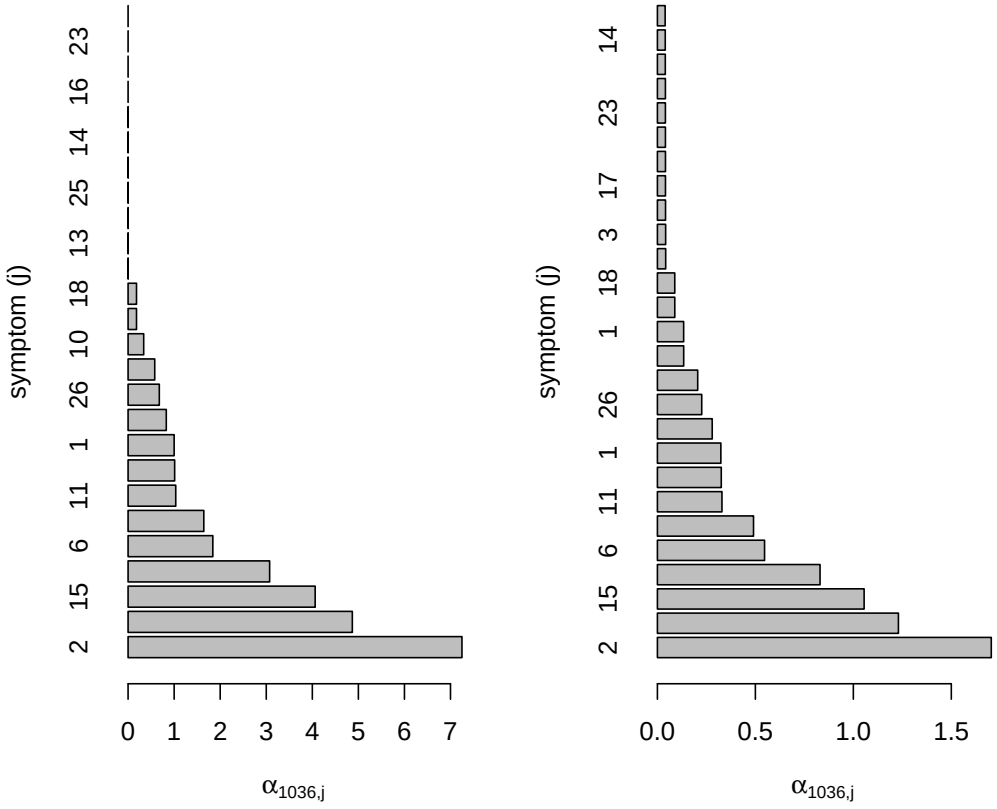


Figure 3.12: Estimates of $\alpha_{1036,j}$ obtained with the maximum likelihood approach (left) and posterior estimates of $\alpha_{1036,j}$ obtained from Bayesian approach (right).

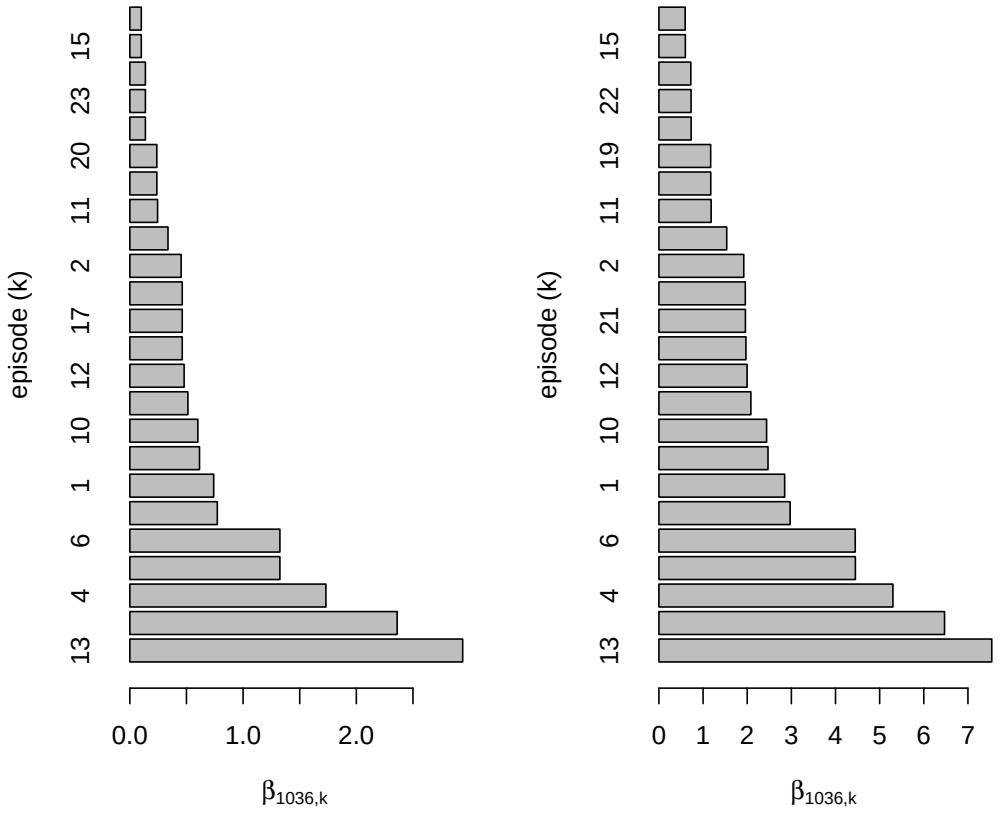


Figure 3.13: Estimates of $\beta_{1036,k}$ obtained with the maximum likelihood approach (left) and posterior estimates of $\beta_{1036,k}$ obtained from Bayesian approach (right).

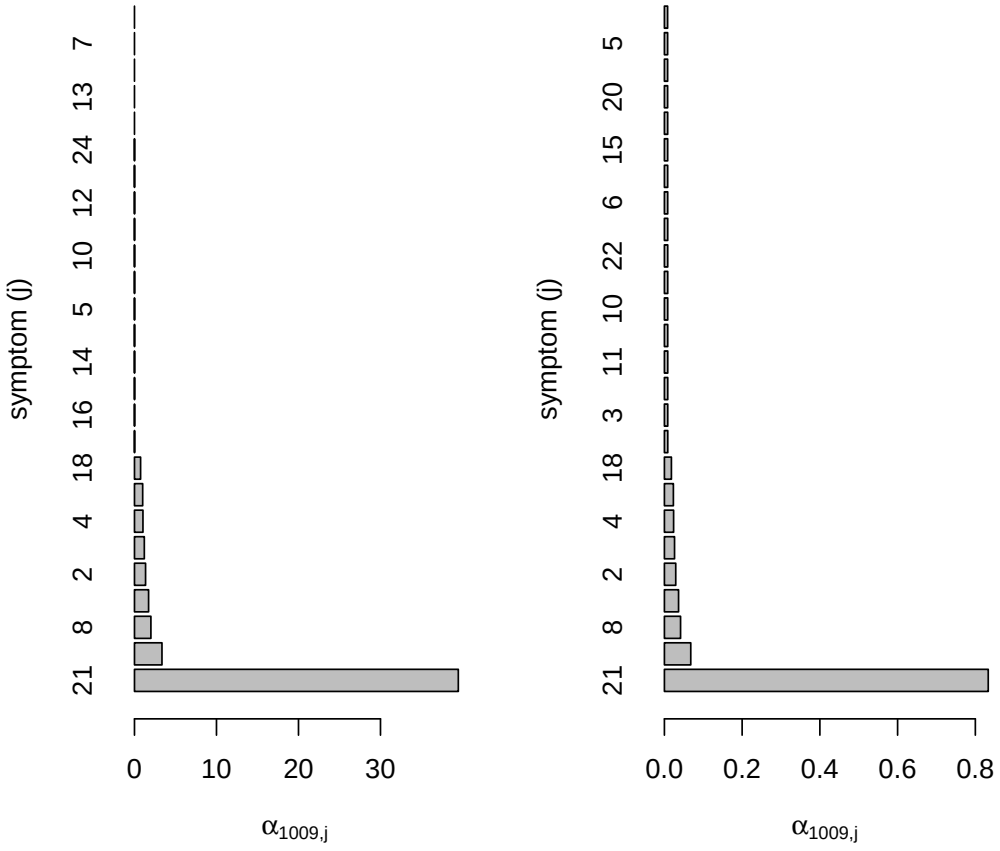


Figure 3.14: Estimates of $\alpha_{1009,j}$ obtained with the maximum likelihood approach (left) and posterior estimates of $\alpha_{1009,j}$ obtained from Bayesian approach (right).

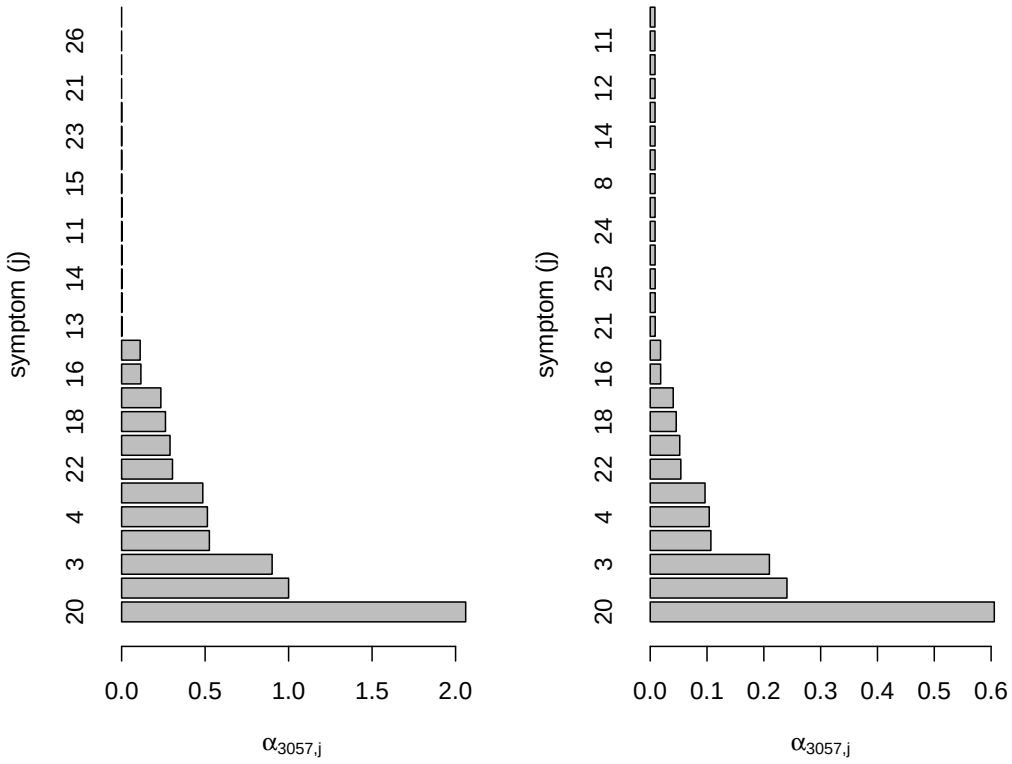


Figure 3.15: Estimates of $\alpha_{3057,j}$ obtained with the maximum likelihood approach (left) and posterior estimates of $\alpha_{3057,j}$ obtained from Bayesian approach (right).

Group	Symptoms
1. Autonomic	sweating, pounding heart, shivering, hunger, thirst, anxiety, tingling and trembling
2. Neuroglycopenic	confusion, drowsiness, weakness, dizziness, difficulty speaking, impaired concentration, unsteady, double vision, tiredness and blurred vision
3. Autonomic/Neuroglycopenic	irritability and feeling warm
4. General malaise	nausea, headache and malaise
5. Other	non-specific awareness and other symptom
6. No symptom	no specified symptom

Table 3.8: The group categorisation of symptoms.

3.4 Grouped symptoms model

The core model framework for the intra-individual consistency (Section 2.5) allows us to expand this model by having an additional source of variation in the precision parameter, σ_i^{-2}. This is achieved by categorising all the symptoms into six distinct groups, four of which are discussed in Chapter 1 (autonomic, neuroglycopenic, general malaise, and autonomic/neuroglycopenic). The additional two groups are "other symptom" for symptoms that were not specified by patients in the report form; and "no symptom" when patients reported hypoglycaemia without experiencing any particular symptom. This group categorisation (Table 3.8) provides us with an additional source of variation that may arise from inherent differences between symptoms in different groups.

The model is previously defined in (2.5.1)-(2.5.5). To allow for group effects, the change made to this model relates to the prior for α_{ij} (2.5.4) which corresponds to propensity of symptom j for patient i. Now, with each symptom being assigned to a specific group, we have α_{ijl}, where l = 1,...,6 indicates group and we assume the following hierarchical prior:

$$\alpha_{ijl} \sim Gamma(\theta, \theta/u_l) \ , \ l = 1,...,6$$

$$\text{giving } \mathrm{E}(\alpha_{ijl}) = u_l, \ l = 1,...,6$$

$$\text{and } \mathrm{Var}(\alpha_{ijl}) = \frac{u_l^2}{\theta}, \ l = 1,...,6, \tag{3.4.1}$$

with

$$\theta \sim Gamma(a_\theta, b_\theta),$$

$$u_l \sim Gamma(a_{u_l}, b_{u_l}). \tag{3.4.2}$$

Given the likelihood function in (2.5.6), the joint posterior distribution can be written as (2.5.7), where the priors for β_{ik} and σ_i have been described in (3.1.1).
The full conditional distributions for parameter α_{ijl} now become

$$\pi(\alpha_{ijl}|\alpha_{\setminus ijl}, \beta_{ik}, \sigma_i, y_{ijk}) \propto \prod_k \Phi\left\{\frac{\log(\alpha_{ijl}\beta_{ik})}{\sigma_i}\right\}^{y_{ijk}} \left[1 - \Phi\left\{\frac{\log(\alpha_{ijl}\beta_{ik})}{\sigma_i}\right\}\right]^{1-y_{ijk}}$$

$$\times \alpha_{ijl}^{\theta-1} \exp(-\frac{\theta}{u_l}\alpha_{ijl}) \text{ for } i = 1,...,I, j = 1,...,J, l = 1,...,6.$$

Here, we set θ equal to 1 since the information in the data at this level of hierarchy is limited, and also to facilitate the convergence of the MCMC algorithm. This effectively corresponds to an exponential prior with mean u_l for α_{ijl}. For similar reasons, we also use $a_{u_l} = b_{u_l} = 1$. Therefore (3.4.1) and (3.4.2) become:

$$\alpha_{ijl} \sim Gamma(1, 1/u_l) \ , \ l = 1,...,6$$

$$u_l \sim Gamma(1, 1) \ l = 1,...,6.$$

Using WinBUGS, we run MCMC for 16,000 iterations for estimating parameters in the grouped symptoms model. The first 1000 iterations were discarded as the burn-in period. The sample paths of the simulation and summary statistics of the posterior

estimates of the group parameter, u_l are provided in Figure 3.16 and Table 3.9 respectively. According to the plots we can see that the chains converge well for all groups. Their densities' plot (Figure 3.18) shows that the mean propensity for each group is different. The mean of propensity for group Autonomic, u_1, is the highest followed by group Neuroglycopenic, u_2. This is consistent with previous findings, where these two groups are equally prominent in identifying the onset of hypoglycaemia (Deary et al., 2001). The "No symptom" group, u_6, shows the highest variability because this group is less reported by patients in this study.

The average mean of symptoms reporting consistency, $\tilde{c}_i$, when using this model is 47.29 with standard deviation of 16.645. The consistency estimates range from 7.20 (Subject 3043) to 95.54 (Subject 1028) respectively. A list of posterior consistency, $\widetilde{c}_i$, with its 95% Bayesian interval for all subjects is provided in Table 3.10.

Figure 3.19 shows the ranking of patient's consistency using the model with grouped symptoms, versus their ranking using the model without grouped symptoms. Patients are ranked in descending order based on their consistency estimates starting from the highest estimates to the lowest. When comparing these estimates with the consistency estimates of the 66 patients in the model without grouped symptoms (Section 3.1), a few patients have remarkably lower or higher consistency estimates values, $\tilde{c}_i$, in this model.

The biggest changes in ranking is Subject 6065, where this patient's consistency was initially the 28^{th} highest of all 66 patients but then the consistency decrease and become the second lowest. Using the grouped symptoms model, Subjects 5009 and 6002 both increase in their consistency estimates. They were ranked at number 63 and 46 before, implying low consistency but climb to higher ranking, 34 and 16 respectively. Another two patients showing big changes in their ranking are patients 4008 and 6018. They were originally at the 56^{th} and 60^{th} rank when using the core model, showing very low consistency but increase to rank number 28 and 41 respectively after we group the symptoms.

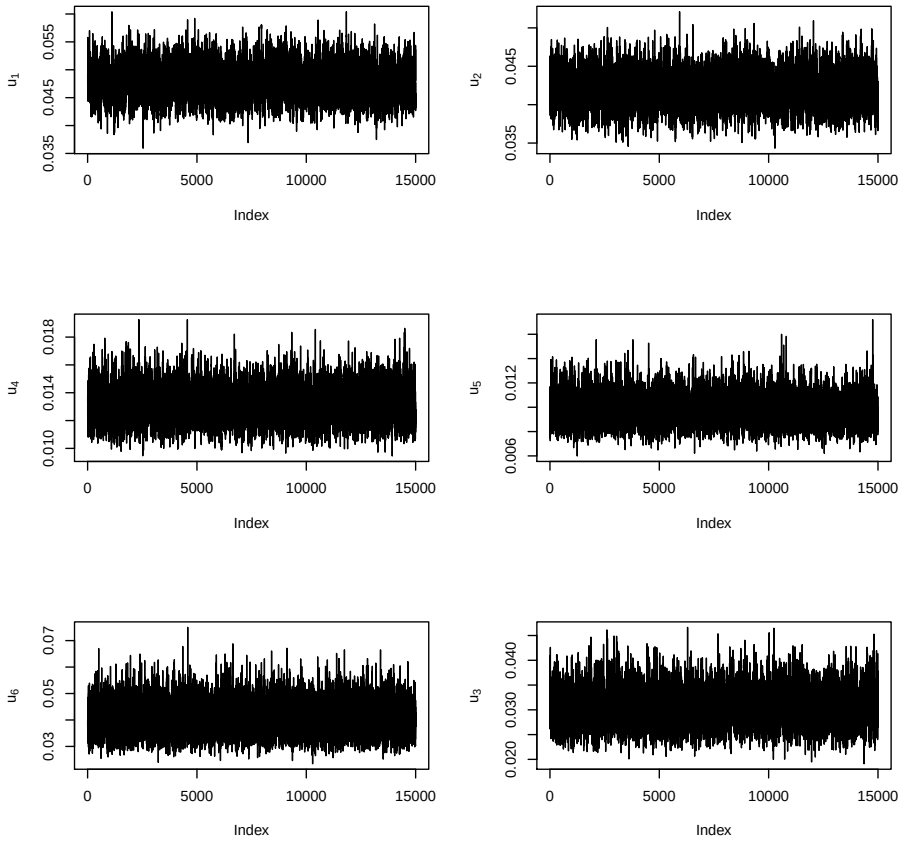


Figure 3.16: Trace plots of group parameter, u_l for $l = 1,...6$ where u_1 refers to Autonomic, u_2 to Neuroglycopenic, u_3 to Autonomic/Neuroglycopenic, u_4 to General malaise, u_5 to Other and u_6 to No symptom.

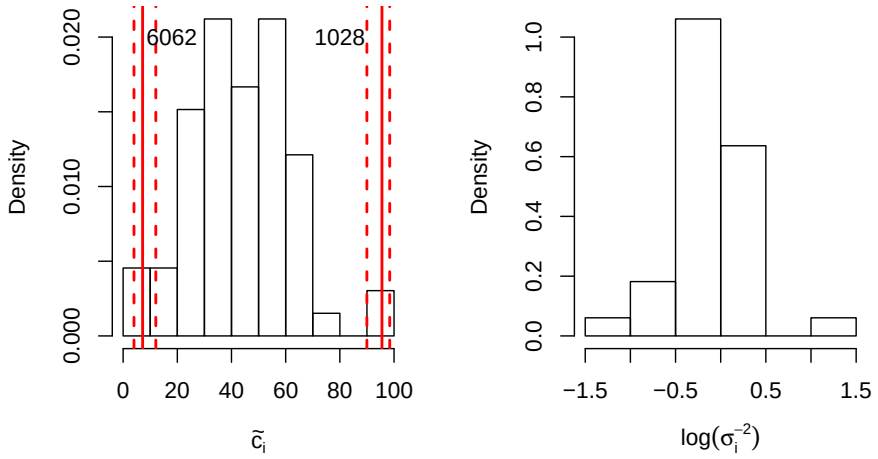


Figure 3.17: Histograms of estimated consistency parameter, $\tilde{c}_i$ and the log estimated precision parameter, σ_i^{-2} for model with grouped symptoms.

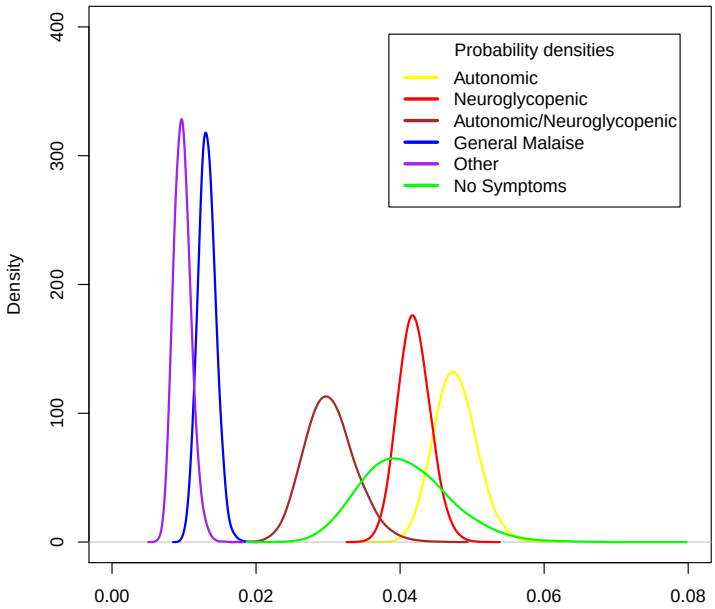


Figure 3.18: Distributions of posterior group propensity, u_l.

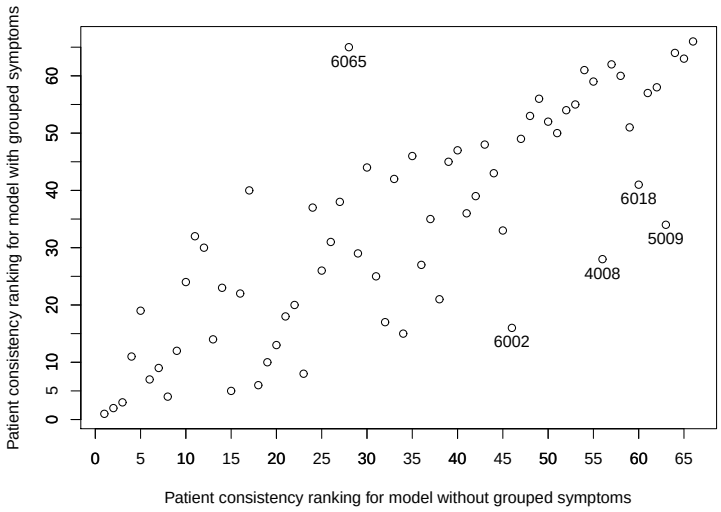


Figure 3.19: Ranking of consistency estimates, $\tilde{c}_i$, of 66 patients for core model (without grouped symptoms), versus model with grouped symptoms.

Group	Mean	SD	2.5% CI	97.5% CI
Autonomic	0.04767	0.00293	0.04227	0.05376
Neuroglycopenic	0.04196	0.00223	0.03780	0.04652
Autonomic/Neuroglycopenic	0.03033	0.00352	0.02417	0.03789
General Malaise	0.01320	0.00123	0.01092	0.01570
Other	0.00979	0.00120	0.00769	0.01236
No Symptom	0.04071	0.00604	0.03033	0.05383

Table 3.9: Posterior estimates of group propensity, u_l.

Group, Subject	No of hypos	$\tilde{c}_i$	SD	95% interval
1008	95	27.84	3.433	(21.60, 35.17)
1009	74	53.40	5.312	(42.94, 63.87)
1015	27	32.09	6.329	(20.79, 46.03)
1021	43	44.08	6.150	(32.64, 56.61)
1025	25	42.90	7.731	(28.64, 58.92)
1028	42	95.37	2.390	(89.23, 98.38)
1036	24	59.07	7.496	(44.11, 73.01)
1039	43	31.77	5.368	(22.02, 42.87)
1055	27	56.63	7.883	(41.02, 71.88)
1057	11	11.61	5.123	(4.93, 23.87)
1086	67	31.51	4.418	(23.34, 40.67)
2009	86	26.58	3.518	(20.15, 33.91)
2010	89	26.71	3.621	(20.14, 34.29)
2012	37	28.99	5.329	(19.59, 40.34)
2013	138	16.15	2.060	(12.43, 20.59)
2015	32	34.55	5.830	(24.24, 46.73)
2021	24	48.99	7.661	(34.06, 64.07)
2022	36	37.94	5.623	(27.40, 49.23)
2027	22	48.49	8.201	(32.97, 65.06)
3001	31	49.33	7.218	(35.39, 63.83)
3015	19	35.80	7.001	(23.19, 50.61)
3016	22	62.23	7.850	(46.39, 76.89)
3022	20	52.79	8.698	(35.85, 69.88)
3024	27	50.78	8.147	(35.06, 66.94)
3029	69	56.59	7.515	(41.20, 71.00)
3043	17	7.18	5.524	(4.11, 11.90)
3046	25	49.13	7.857	(34.01,64.45)
3048	31	59.95	7.101	(45.74, 73.28)
3050	44	55.88	5.889	(44.31, 67.14)
3052	105	51.24	4.748	(41.96, 60.58)
3057	51	34.64	5.182	(25.03, 45.47)
3065	26	36.83	6.696	(24.84, 51.01)
3067	26	47.66	7.271	(33.46, 61.88)
4003	23	63.59	7.277	(48.81, 77.27)
4008	59	50.57	5.233	(40.68, 61.13)
4013	39	74.45	7.192	(59.29, 86.94)
4023	134	26.06	3.109	(20.31, 32.48)
4028	12	10.84	3.430	(5.67, 19.08)
4032	10	32.47	9.572	(17.49, 54.43)
4034	55	65.75	5.245	(55.27, 75.63)
4043	35	35.45	5.781	(24.65, 47.28)
4045	38	35.68	6.048	(24.67, 47.93)
4049	22	56.17	7.687	(41.31, 71.26)
4061	12	7.612	3.938	(2.92, 18.16)
4063	42	29.07	5.000	(20.16, 39.37)
4072	39	42.85	6.432	(30.70, 55.95)
4076	29	40.75	6.662	(28.42, 54.16)
5004	26	44.53	7.422	(31.47, 58.51)
5009	36	48.98	6.411	(36.78, 61.65)
5023	20	52.03	8.594	(35.32, 69.26)
5026	24	65.06	7.550	(49.87, 79.12)
5029	45	61.13	6.021	(48.68, 72.37)
5044	70	32.23	4.290	(24.20, 41.11)
5045	68	42.31	5.261	(32.49, 53.20)
5088	87	19.77	2.929	(14.39, 25.75)
6002	34	56.91	6.438	(44.39, 69.60)
6010	19	54.81	8.210	(38.94, 70.69)
6018	39	42.43	5.854	(31.55, 54.30)
6019	64	32.08	4.462	(23.86, 41.32)
6023	59	28.16	4.390	(21.64, 37.15)
6038	79	28.32	3.850	(21.25, 36.30)
6056	28	64.62	6.959	(50.25, 77.16)
6058	20	68.63	7.418	(53.20, 82.14)
6062	13	9.94	3.061	(5.29, 17.31)
6064	17	92.34	3.378	(84.20, 96.97)
6065	39	17.43	4.516	(10.44, 27.98)

Table 3.10: Posterior estimates of consistency (Mean, standard deviation, and 95% Bayesian interval) for all subjects with grouped symptoms model.

3.5 Model Assessment

In Section 3.4, we introduce the grouped symptoms models as one of our consistency estimation models. In this section, we would like to test if this model is adequate. To do this, we utilise the concept of stochastic latent residuals, z_{ijk}, described below.

First, we test the grouped symptoms model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Therefore, patient i reports symptom j at episode k when $\tau_{ijk} \leq \alpha_{ij}\beta_{ik}$. Thus, the observed variable, Y_{ijk} is equal to 1 when symptom j is reported at episode k by patient i. Otherwise, Y_{ijk} takes value zero.

We now consider the stochastic latent residuals, z_{ijk}, that would give rise to the observed data under the considered model. Following the concept of generalised residuals (Cox and Snell, 1968), the data can be regarded as generated through a functional model, $g(\cdot)$, (Dawid and Stone, 1982) depending on the vector of all model parameters and latent variables, say $\boldsymbol{\theta}$, i.e.

$$\boldsymbol{y} = g_{\boldsymbol{\theta}}(\boldsymbol{z}) \tag{3.5.1}$$

where $\boldsymbol{z} \sim U(0,1)$ are generalised residuals. Then, in the general case, the function in (3.5.1) can be inverted to give the stochastic latent residuals

$$\boldsymbol{z} = g_{\boldsymbol{\theta}}^{-1}(\boldsymbol{y}).$$

For the assumed discrete model we have

$$y_{ijk} = I\{z_{ijk} \leq p_{ijk}\}$$

where $z_{ijk} \sim U(0,1)$ and $I\{\cdot\}$ is the indicator function. This implies that, under the assumed model, we can write

$$z_{ijk} = y_{ijk}u_1 + (1 - y_{ijk})u_2$$

where $u_1 \sim U(0, p_{ijk})$ and $u_2 \sim U(p_{ijk}, 1)$. Therefore, if the model is adequate

$z_{ijk} \sim U(0,1)$ and we can obtain a p-value for testing the hypothesis of this uniform distribution. To implement this method, we run 10,000 MCMC iterations for this model and obtain the latent residual, z_{ijk}, for each subject such that

$$\text{If } \begin{cases} Y_{ijk} = 1, & z_{ijk}^{(1)} \sim U(0, \hat{p}_{ijk}) \\ Y_{ijk} = 0, & z_{ijk}^{(0)} \sim U(\hat{p}_{ijk}, 1) \end{cases}$$

where $\hat{p}_{ijk}$ is the estimated probability of patient i reporting symptom j at episode k at each iteration. Therefore, if the tested hypothesis is correct,

$$z_{ijk} = (z_{ijk}^{(1)}, z_{ijk}^{(0)}) \sim U(0,1).$$

We run a Kolmogorov-Smirnov goodness-of-fit test on each posterior sample of residuals obtained in each MCMC iteration, and calculate a corresponding p-value, π_γ, where $\gamma = 1, 2, 3, ..., 10,000$ iterations. This will give a posterior distribution $f(\pi|Y_{ijk})$ where Y_{ijk} denotes the observation data. As preliminary checking, we observe the histogram of the residuals $\boldsymbol{z}$ for each patient i. We only show here the histogram for one patient, Subject 4028 in Figure 3.20. The distribution pattern suggests that the residuals do follow a Uniform(0,1) distribution. To further confirm the distribution of $\boldsymbol{z}$ we also check on the histogram of p-values for this patient, $\pi_{(4028)}$. This is given in Figure 3.21. From this histogram, we can say there is no evidence against the adequacy of fit of our model.

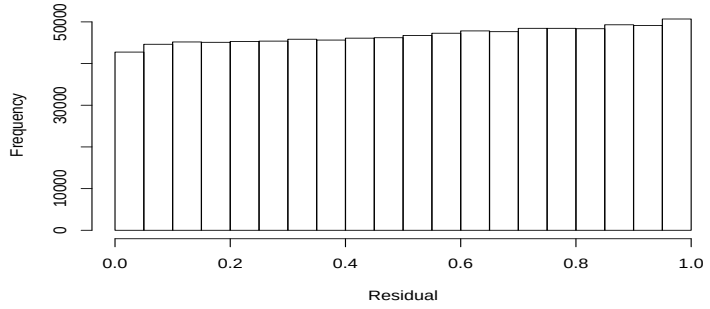


Figure 3.20: Histogram of stochastic latent residuals for model with grouped symptoms using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ for patient 4028.

Since threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$ is never favoured in Section 3.3.1, we omit that threshold here. Therefore, only results using thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$

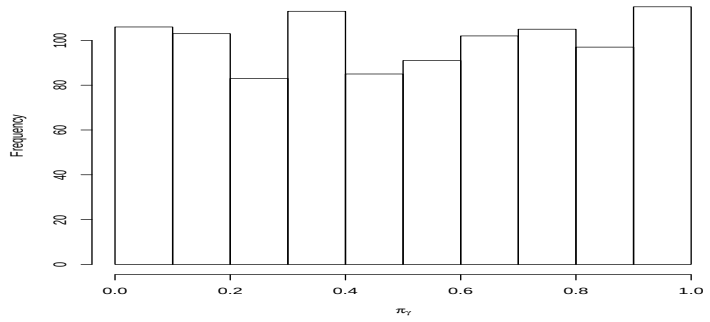


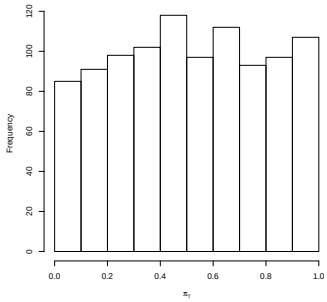
Figure 3.21: Posterior distribution of p-values, π, for fit of model with grouped symptoms using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ for patient 4028.

and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ are shown in Figure 3.22. We present here the histograms of p-values, π_γ, for four patients when using the two thresholds. Observing the posterior distributions of π_γ for subjects 5088, 4008, 2013, and 4023, once again we can see that there is no strong evidence against the models tested, although it appears that there is more evidence against the model when the model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ is fitted. This can be seen from the higher concentration of p-values close to zero in two of these four patients.

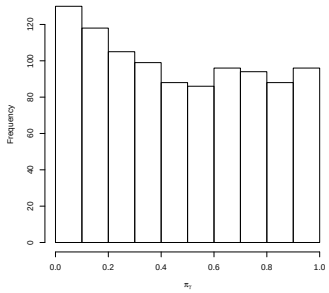
As implied earlier, to have strong evidence against a tested model, i.e. to reject the hypothesis that the model is adequate, the posterior p-values, π_γ, should be very small. Therefore as a measure of model goodness of fit, we calculate the proportion of π_γ less than 0.05, $\Pr(\pi_\gamma < 0.05)$, for each subject in the models discussed here. For comparison purposes, we can say that cases with greater $\Pr(\pi_\gamma < 0.05)$ show stronger evidence against the model fit.

Bar plots in Figure 3.23 display the $\Pr(\pi_\gamma < 0.05)$ obtained from the grouped symptoms model when using different thresholds. For 67% of the 66 subjects the proportions of $\pi_\gamma < 0.05$ suggest that better fit of the model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. For these patients, their $\Pr(\pi_\gamma < 0.05)$ when using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$ (yellow bars) are higher than when using $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ which is indicated by the red bars.

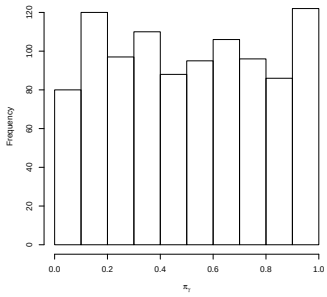
We also repeat the same exercise for comparing the models with and without grouped



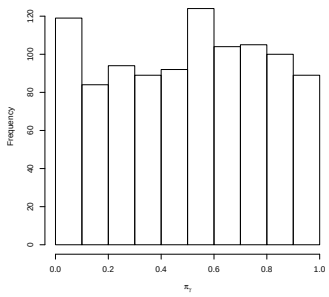
(a) Patient 5088, $\alpha_{ij}\beta_{ik}$



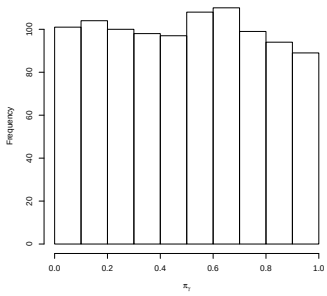
(b) Patient 5088, $\alpha_{ij} + \beta_{ik}$



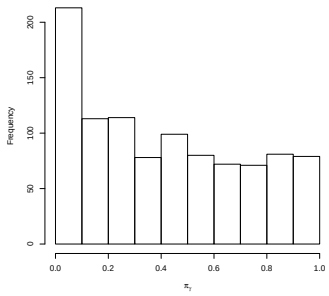
(c) Patient 4008, $\alpha_{ij}\beta_{ik}$



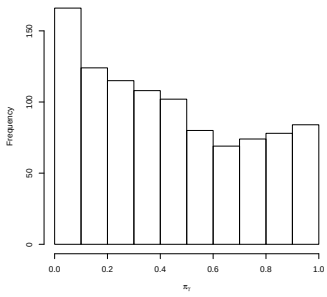
(d) Patient 4008, $\alpha_{ij} + \beta_{ik}$



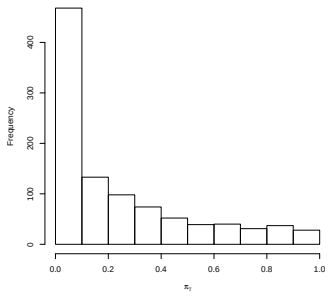
(e) Patient 2013, $\alpha_{ij}\beta_{ik}$



(f) Patient 2013, $\alpha_{ij} + \beta_{ik}$



(g) Patient 4023, $\alpha_{ij}\beta_{ik}$



(h) Patient 4023, $\alpha_{ij} + \beta_{ik}$

Figure 3.22: Posterior distribution of p-values, π, for fit of model with different thresholds for patients 5088, 4008, 2013 and 4023 in grouped symptoms model.

symptoms. For both models the $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ threshold is used, and the proportion $\Pr(\pi_\gamma < 0.05)$ are calculated and plotted in Figure 3.24. Only seven patients show higher $\Pr(\pi_\gamma < 0.05)$ when the grouped symptoms model is used (indicated by yellow bars), compared to the model without grouped symptoms (indicated by red bars). These subjects are 1055, 2010, 2015, 4013, 4049, 4063, and 4064. $\Pr(\pi_\gamma < 0.05)$ calculated from the grouped symptoms model and the without grouped symptoms model for patient 2010 are 0.227 and 0.082 respectively. Patient 2015 has $\Pr(\pi_\gamma < 0.05) = 0.096$ when using the grouped symptoms model and $\Pr(\pi_\gamma < 0.05) = 0.052$ when using the without grouped symptoms model. This suggests that the model with grouped symptoms fits our data better.

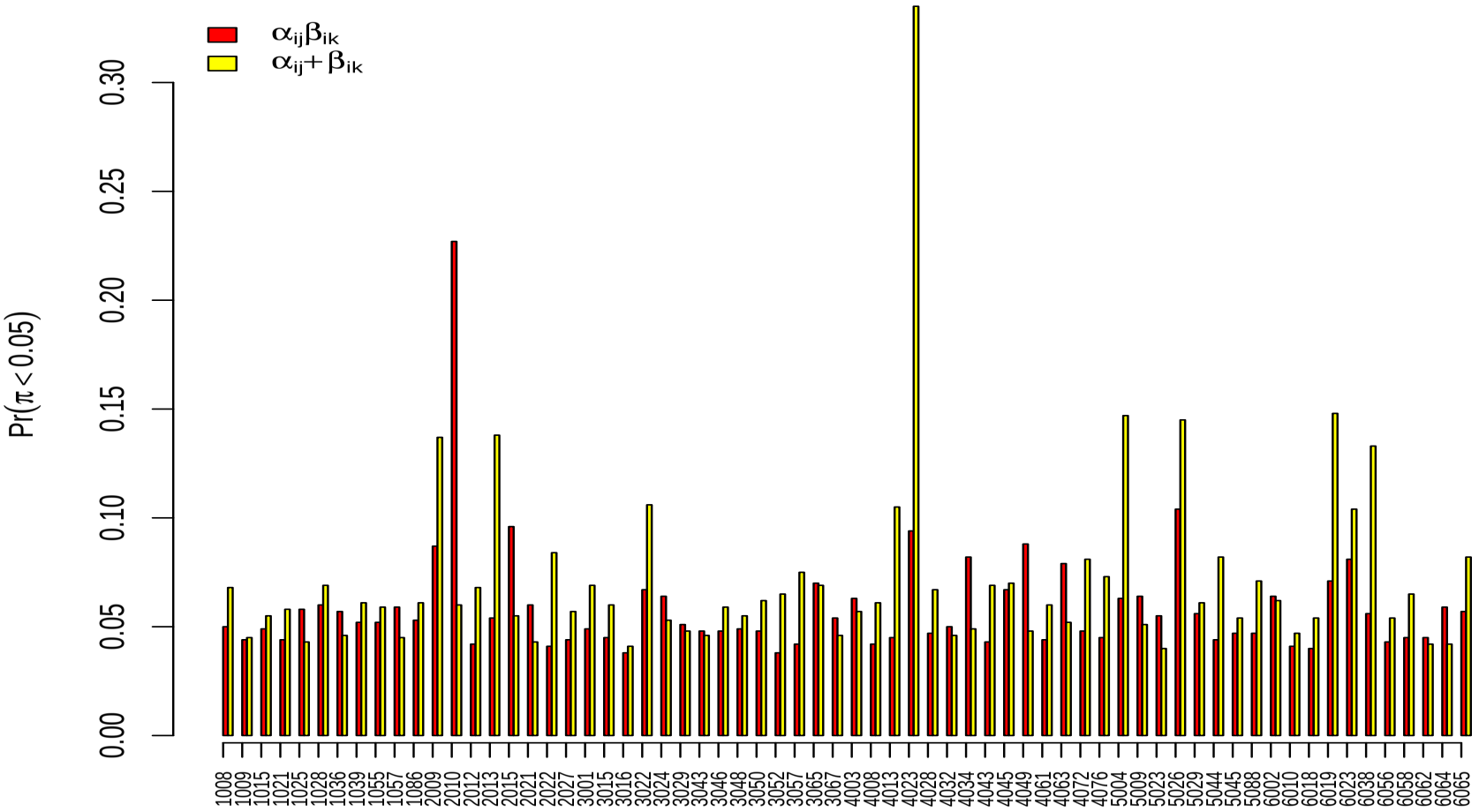


Figure 3.23: Bar plots comparing the proportion of p-values < 0.05 between different thresholds when using the grouped symptoms model.

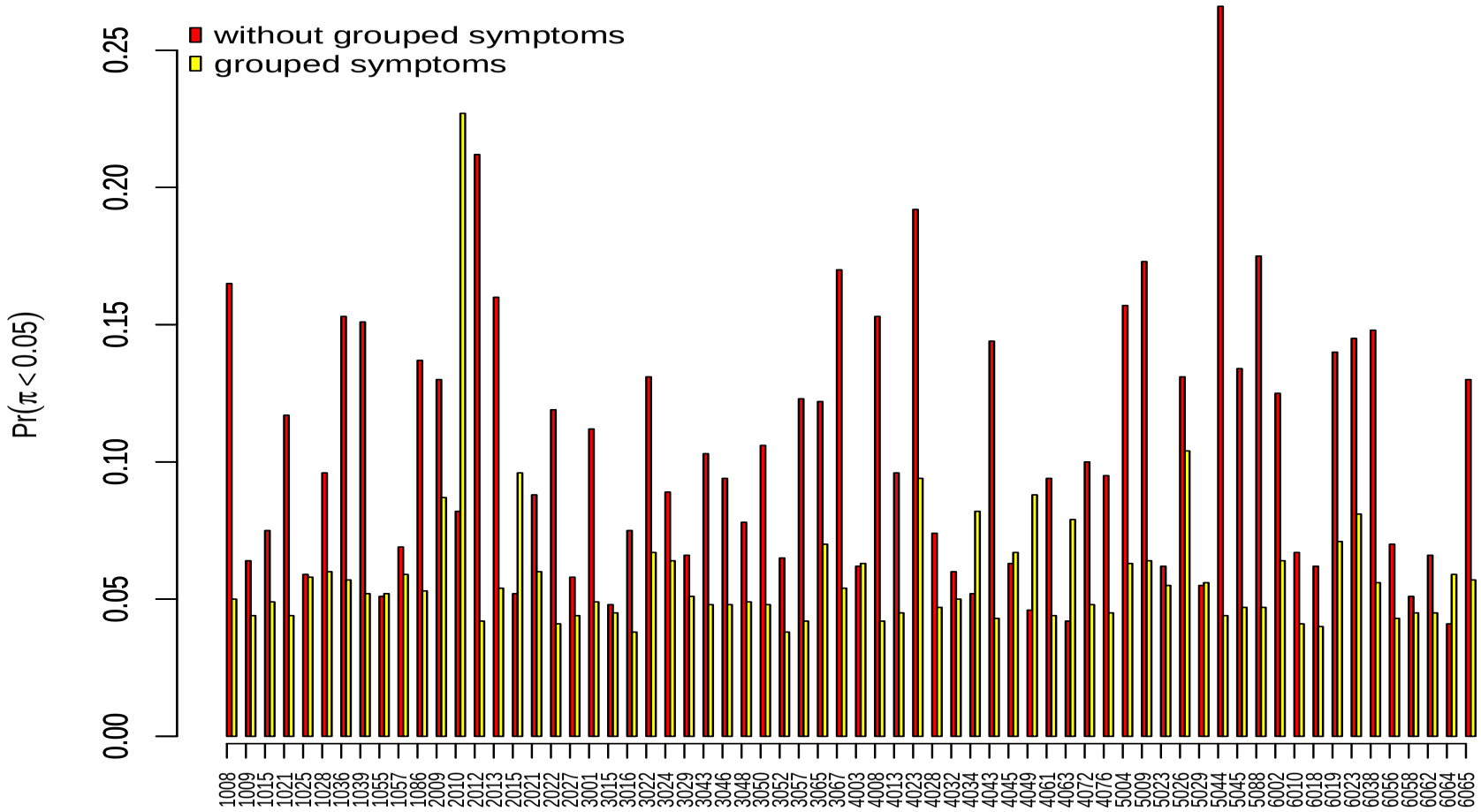


Figure 3.24: Bar plots comparing the proportion of p-values < 0.05 between different models when using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$.

3.6 Model verification and posterior predictive checking

To examine our model's predictive capability, we run a validation approach relying on the posterior predictive distribution. This approach is commonly used for checking the model's suitability, and is based on work that was elaborated by Rubin (1984) and later expanded by Gelman et al. (1996). The purpose of the analysis is to compare the observed data with values predicted from our model. Also, one of the important aims of this work is to develop a model which can be used to make prediction of values of interest with quantified confidence. Later in this thesis (Chapter 6), given information on some specific characteristics about a patient, we will aim to predict how consistent the patient is in reporting hypoglycaemia. This can be used in order to assist early detection of hypoglycaemia and give necessary advice to the patient.

Our observations, Y_{ijk} are binary data that take value 1 if patient i reported symptom j in episode k value zero otherwise. We define $Y_{ijk}^{(p)}$ as the predicted data, such that these are the data that will be obtained if we use the same model to do prediction. We start off by selecting randomly 10% of the total number of observations we have, which are then used as the validation sample. Then, we fit the examined model to the remaining data. The fitted model is subsequently used to do prediction on symptom reporting for the sampled patients episodes, $Y_{ijk}^{(p)}$ for i,j,k in the sample. Recall from (2.5.1), Y_{ijk} is Bernoulli distributed with probability, p_{ijk}. We sample the posterior probability, p_{ijk}, from the fitted model and use it to obtain the reporting prediction, $Y_{ijk}^{(p)}$. Consequently, we compare the total number of predicted reportings, N_p to the total number of observed reportings, N_{obs}. Accordingly, we compare the distributions of Y_{ijk} and $Y_{ijk}^{(p)}$.

Streftaris et al. (2013) also use another four measures to assess and describe the usefulness of their model predictions. The measures are related to sensitivity, specificity and predictive values. Here, sensitivity is defined as the proportion of experienced symptoms that are correctly predicted as being reported by the models whereas specificity is the proportion of symptoms that have not been experienced which are

correctly predicted as not reported by the model. Ideally, a good predictive model should have high sensitivity and specificity. However these two measures are often inversely proportional, meaning as sensitivity increases specificity decreases and vice versa. To evaluate the probability of the model giving correct prediction we also calculate the positive predictive value (PPV) and negative predictive value (NPV). These four measures were calculated using Y_{ijk} and $Y_{ijk}^{(p)}$ in the validation sample and are defined as follows.

a)

$$\text{PPV} = \frac{\sum_{ijk} Y_{ijk} Y_{ijk}^{(p)}}{\sum_{ijk} Y_{ijk}^{(p)}} \qquad \text{for } i, j, k \text{ in the sample.}$$

Therefore PPV is the proportion of symptoms with positive prediction that were correctly classified as reported. PPV measures the probability of patient i truly experiencing symptom j at episode k given that the model predicts the symptom is likely to be experienced.

b)

$$\text{NPV} = \frac{\sum_{ijk}(1 - Y_{ijk})(1 - Y_{ijk}^{(p)})}{\sum_{ijk}(1 - Y_{ijk}^{(p)})} \qquad \text{for } i, j, k \text{ in the sample.}$$

NPV is the proportion of symptoms with negative reporting prediction that were correctly classified as absent. NPV measures the chance of patient i having symptom j not present at episode k given that the model predicts that it is not likely to be reported.

c)

$$\text{TPR} = \frac{\sum_{ijk} Y_{ijk} Y_{ijk}^{(p)}}{\sum_{ijk} Y_{ijk}} \qquad \text{for } i, j, k \text{ in the sample.}$$

True Positive Rate (TPR), also known as the sensitivity of our predictive model, measures the ability of the model to correctly predict if symptom j occurs at episode k.

d)

$$\text{TNR} = \frac{\sum_{ijk}(1 - Y_{ijk})(1 - Y_{ijk}^{(p)})}{\sum_{ijk}(1 - Y_{ijk})} \qquad \text{for } i, j, k \text{ in the sample.}$$

True Negative Rate (TNR), or also called specificity, represents the capacity of the model to predict that symptom j is not reported at episode k when the symptom is truly absent.

3.6.1 Results

We applied the posterior predictive checking approach to study the predictive ability of the core model (without grouped symptoms) and the grouped symptoms model, using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Graphical plots in Figure 3.25 show the posterior distributions of the total predicted number of reporting symptoms, with blue (dotted) lines marking the total number of predicted reportings, N_p, whereas the red (solid) lines refer to the total observed value, N_{obs} for subjects 5004, 6038, 4045, and 5009.

The symptom reportings i.e. the total observed value, N_{obs}, for subject 5004 is 9 and the prediction from the non-grouped symptoms model is $N_p = 12.19$ with 95% CI (7, 19), while the prediction made by the grouped symptoms model is much closer, $N_p = 10.14$ with 95% CI (5,16). Prediction for patient 6038 is good when using both grouped and non-grouped symptoms model as we can see the red lines ($N_{obs} = 21$) lie within the predicted number of reporting symptoms' distributions from both models.

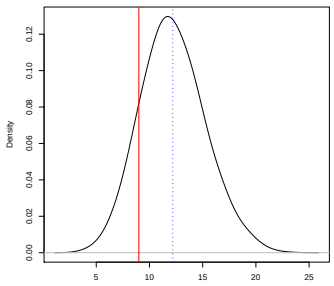
The reporting symptoms for patient 4045 is very well predicted by the grouped symptoms model as indicated by the blue and red lines that almost overlap. The prediction made was 15.26 with 95% CI (9,22) and the observed value is 15. However, the non-grouped symptoms model also made a good prediction, although it is slightly over estimated ($N_p = 17.63$). The posterior predictive checking for subject 5009 also shows that the grouped symptoms model yield a slightly better prediction in comparison to the non-grouped symptoms model. The predicted value from the non-grouped model, $N_p = 31.45$ lies at the lower tail of the estimated distribution (95% CI (24,39)) while

the grouped-symptoms model gives $N_p = 27.18$ and the 95% CI (20, 35) comfortably contains $N_{obs} = 24$.

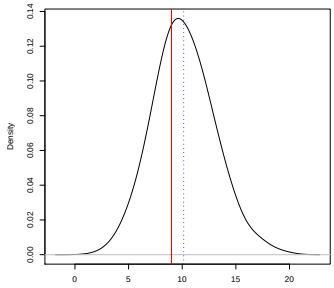
We also test the performance of different thresholds with the core model. Figure 3.26 gives the posterior distributions of the total predicted number of symptom reportings for subjects 5009, 3057, 2015 and 6038. With each threshold, the predicted distributions comfortably contain the total number of observation, N_{obs} (represented by red solid lines). This indicates that for these patients, we cannot distinguish between the three threshold models in terms of their predictive ability.

Finally, we want to compare the performance of prediction for different models when all patients are used in the analysis. We omit threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$ here since this threshold did not perform very well using other criteria in earlier analysis (see Section 3.3.1). Therefore, we test the models with and without grouped symptoms using thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$. The results are provided in Table 3.11 and the posterior distributions of the predicted number of symptoms reported are displayed in Figure 3.27. The plots show that among the four models, the model with grouped symptoms with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ gives the closest predicted value, N_p, to the observed number of symptoms reported. The total number of symptoms predicted to be reported, is 754.3, with a 95% credible interval of (713,797), which contains the observed number of reported symptoms, 771. The other three models considered here do not perform well in terms of this prediction, with the corresponding posterior predictive distributions failing to contain the true value.

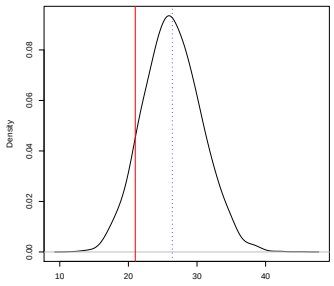
Regarding the other four predictive measures presented in Table 3.11, we notice that the models explored here do not display substantial differences. It is also obvious that prediction referring to symptoms not being experienced (NPV, TNR) is much more successful, as compared to prediction for reported symptoms (PPV, TPR). The fact that the developed models perform better in predicting that symptoms will not be reported, may be explained by the nature of the data, where the frequency of reporting symptoms is relatively low (771/7033). The proportion of symptoms with



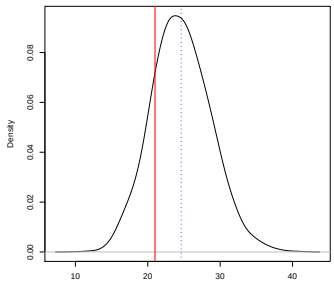
(a) N_p=12.19, N_{obs}=9



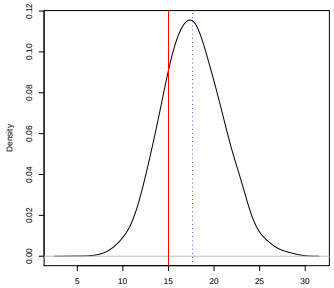
(b) N_p=10.14, N_{obs}=9



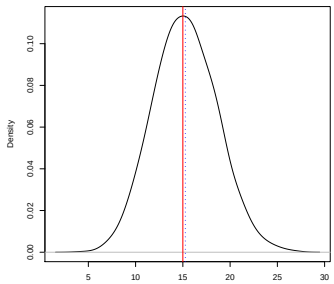
(c) N_p=26.39, N_{obs}=21



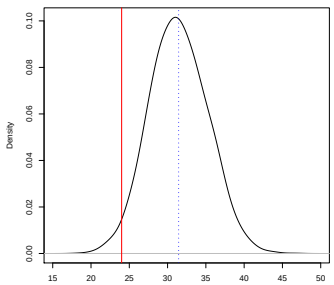
(d) N_p=24.61, N_{obs}=21



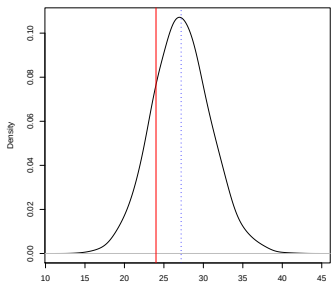
(e) N_p=17.63, N_{obs}=15



(f) N_p=15.26, N_{obs}=15

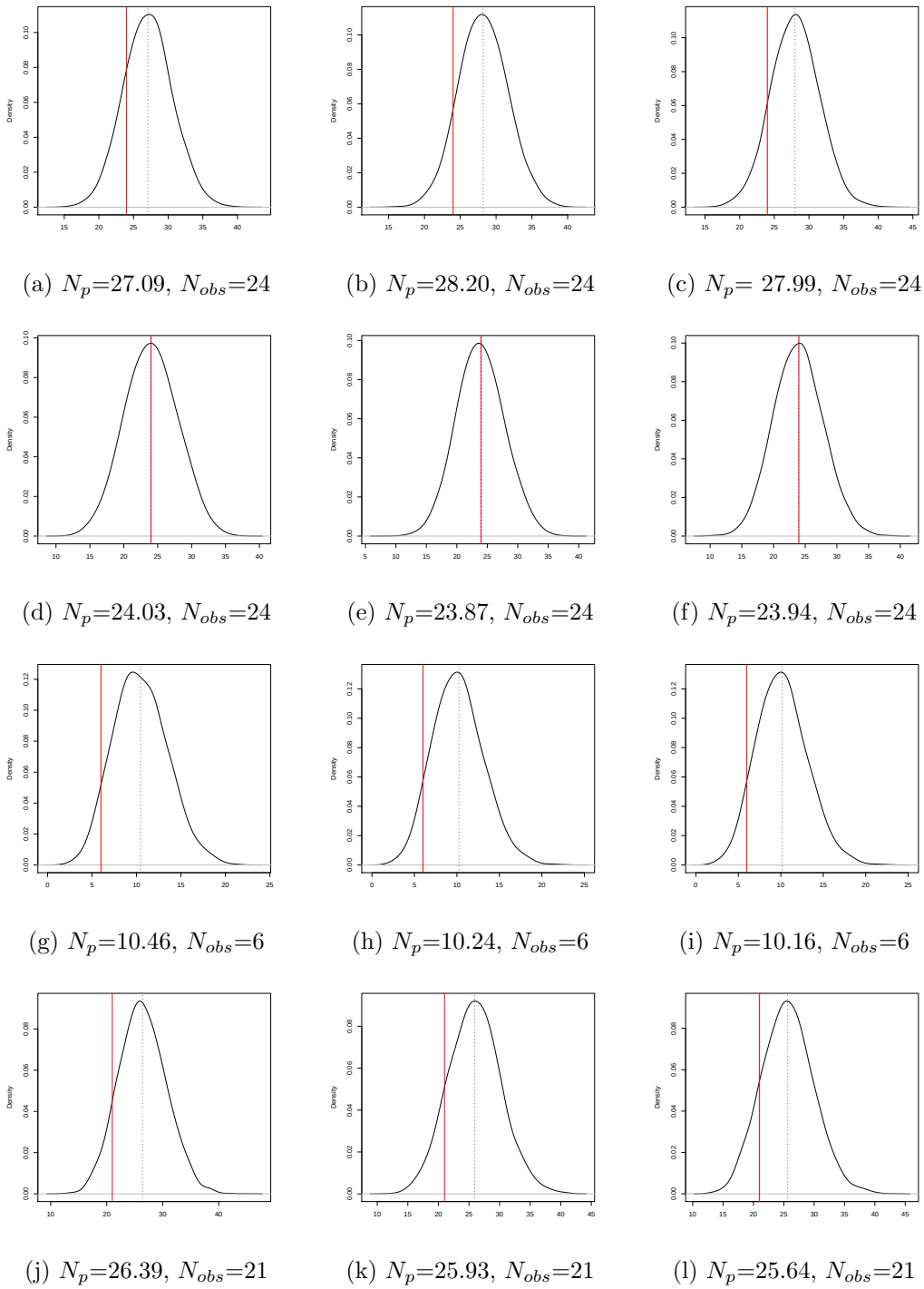


(g) N_p=31.45, N_{obs}=24



(h) N_p=27.18, N_{obs}=24

Figure 3.25: Posterior density plots of number of reportings, N_p, for patients 5004, 6038, 4045, and 5009 under non-grouped symptoms model (left) and grouped symptoms model (right) using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. Blue dotted lines show the number of predicted symptom reportings of each model, and red lines represent the true number of reported symptoms, N_{obs}.



(a) N_p=27.09, N_{obs}=24 (b) N_p=28.20, N_{obs}=24 (c) N_p= 27.99, N_{obs}=24

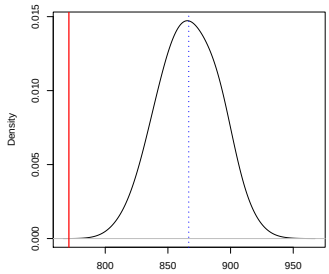
(d) N_p=24.03, N_{obs}=24 (e) N_p=23.87, N_{obs}=24 (f) N_p=23.94, N_{obs}=24

(g) N_p=10.46, N_{obs}=6 (h) N_p=10.24, N_{obs}=6 (i) N_p=10.16, N_{obs}=6

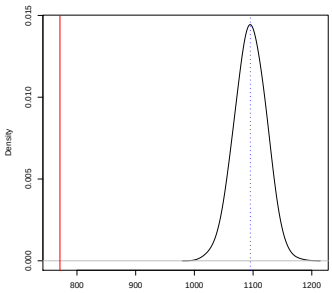
(j) N_p=26.39, N_{obs}=21 (k) N_p=25.93, N_{obs}=21 (l) N_p=25.64, N_{obs}=21

Figure 3.26: Posterior density plots of number of reportings, N_p, for patients 5009, 3057, 2015, and 6038 under non-grouped symptoms model using thresholds $\alpha_{ij}\beta_{ik}$ (left), $\alpha_{ij} + \beta_{ik} + \alpha_{ij}\beta_{ik}$ (middle) and $\alpha_{ij} + \beta_{ik}$ (right). Blue dotted lines show the number of predicted symptom reportings of each model, and red lines represent the true number of reported symptoms, N_{obs}.

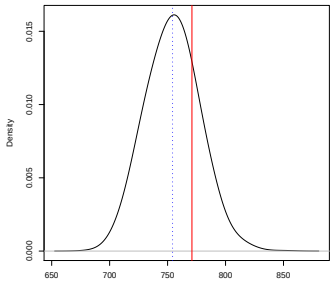
positive prediction that were correctly classified as reported is highest when using the core model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$, i.e. PPV = 0.413, whereas the chance of a symptom not present in an episode given that the model predict it will not be reported is also highest with this model (NPV = 0.933).



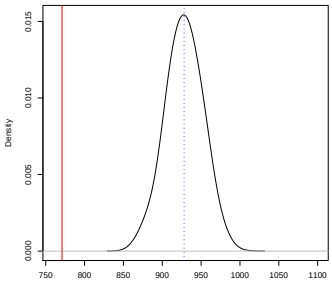
(a) Non-grouped symptoms model, $\alpha_{ij}\beta_{ik}$



(b) Non-grouped symptoms model, $\alpha_{ij} + \beta_{ik}$



(c) Grouped symptoms model, $\alpha_{ij}\beta_{ik}$



(d) Grouped symptoms model, $\alpha_{ij} + \beta_{ik}$

Figure 3.27: Posterior density plots of number of reportings, N_p, for all patients under different models. Blue dotted lines show the number of predicted symptom reportings of each model, and red lines represent the true number of reported symptoms, N_{obs}.

From the analysis in Sections 3.4 and 3.5, we can conclude that among the three thresholds explored, $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ fits the data best. Also the model with grouped symptoms performs better than the model with non-grouped symptoms in predicting symptom reporting.

	$\alpha_{ij}\beta_{ik}$		$\alpha_{ij} + \beta_{ik}$	
	mean	95%CI	mean	95%CI
PPV	0.413	(0.390, 0.436)	0.330	(0.310, 0.348)
NPV	0.933	(0.930, 0.936)	0.931	(0.928, 0.935)
TPR	0.464	(0.435, 0.489)	0.469	(0.437, 0.498)
TNR	0.919	(0.913, 0.925)	0.883	(0.876, 0.890)
N_p	866.626	(820, 908)	1095.216	(1044, 1142)

(a) Non-grouped symptoms model

	$\alpha_{ij}\beta_{ik}$		$\alpha_{ij} + \beta_{ik}$	
	mean	95%CI	mean	95%CI
PPV	0.408	(0.383, 0.432)	0.380	(0.359, 0.402)
NPV	0.926	(0.923, 0.930)	0.931	(0.928, 0.935)
TPR	0.399	(0.368, 0.429)	0.457	(0.431, 0.492)
TNR	0.929	(0.923, 0.934)	0.908	(0.902, 0.914)
N_p	754.300	(713, 797)	928.084	(876, 969)

(b) Grouped symptoms model

Table 3.11: Model predictions for validation sample for all subjects (The number of reported symptoms, $N_{obs} = 771$).

3.7 Concluding remarks

This chapter develops and assesses statistical models for the consistency of hypoglycaemic symptom reporting by individual patients. Initially, we fit models with three different thresholds to the core model. Comparing the deviance information criterion values of the three models, we conclude that the core model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ fits the data better. These findings are also confirmed by using ML estimation.

Next, we introduce an additional source of variation in the consistency model by grouping the symptoms into different groups according to their cause. Thus, variation now comes from within each group and between groups. Output from this model shows that the mean propensity of symptoms in each group is different. Section 3.4 discusses the assessment of models with different thresholds using the concept of stochastic latent residuals. It was found that the grouped symptoms model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ fits the data better. Finally, we perform model verification and posterior predictive checking to verify which model is best in predicting symptom reporting. The grouped symptoms model with threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ has more reliable predictive ability as compared to other models.

Chapter 4

Analysing Episodes Occurring Within 24 Hours From Preceding Episodes

In this chapter, we investigate the effect of adding data from episodes that occurred within 24 hours from their preceding episode. We start by checking correlations between the two cases: excluding episodes within 24 hours and including those episodes. We then compare the intensities of episodes between the two cases. We further our analysis by measuring the relative mean difference in consistency for both situations and also calculate the consistency's coefficient of variation. The comparisons show that adding these new episodes does not have a significant impact on the intensity of episodes. Therefore, in the final section of this chapter we look at the similarity of the posterior distributions of consistency when we include and exclude the episodes within 24 hours.

4.1 Correlation and intensity of hypoglycaemic episodes

In a previous study (Zammitt et al., 2011), any episode occurring within 24 hours from a previous episode was excluded from further consideration because it was argued that it may reduce the intensity of subsequent hypoglycaemic episodes. Such

episodes could potentially affect our model given that it is based on the patient's inherent propensity to report a given symptom and the intensity of that symptom in a given episode.

In this chapter, we are interested to see what happens if these discarded episodes are included in subsequent analyses, as in real life patients would want to know whether their hypoglycaemic episodes occurring soon after a previous episode are more likely to generate inconsistent symptoms. There are 775 episodes which occurred within 24 hours from their previous episode. Before we use them, we need to analyse these episodes to explore if the claim that they have diminished intensity is true. If it is shown that it is not true and symptoms are not blunted, then it is worth including those episodes in our study.

First, we want to compare the correlation between the intensity of episodes occurring within 24 hours, with the correlation of episodes after we take out episodes occurring within 24 hours. Correlation A, r_A, refers to the correlation between the intensity of episodes not containing any episode that occurred within 24 hours of the previous episode, while correlation B, r_B, is the correlation for pairs of episodes reported within 24 hours from their previous episodes. With β_{ik} providing, as previously, the intensity parameter for each episode k for patient i, the correlation coefficient, r, is calculated as:

$$r_i = \frac{\sum_{k=1}^{M}(\beta_{ik} - \overline{\beta}_i)(\beta'_{ik} - \overline{\beta'}_i)}{\sqrt{\sum_{k=1}^{M}(\beta_{ik} - \overline{\beta}_i)^2 \sum_{k=1}^{M}(\beta'_{ik} - \overline{\beta'}_i)^2}}$$

where $i = 1,...,I$ refers to patients and $k = 1,...,M$ to pairs of episodes. For correlation A, r_A, we calculate the correlation for pairs $(\beta_{ik}, \beta'_{ik})$ such that β'_{ik} is the episode following β_{ik} given that β'_{ik} is not reported within 24 hours of β_{ik}. To calculate correlation B, r_B, we pair the episodes $(\beta_{ik}, \beta'_{ik})$ such that β'_{ik} is an episode that occurred within 24 hours from episode β_{ik}. Also, $\overline{\beta_i} = \frac{\sum_{k=1}^{M}\beta_{ik}}{M}$ denotes the mean of episodes' intensity for patient i, over the relevant M pairs of episodes and similarly for $\overline{\beta'}_{ik}$. The computation of r is based on posterior estimates of β_{ik} obtained from the MCMC es-

timation, and therefore estimates of the entire posterior distribution of r are available.

This correlation measures how much two successive episodes are correlated and as usual will give values between -1 and 1. Correlation equal to 1 implies the episodes have perfect positive linear relationship while -1 indicates that the episodes have perfect negative linear relationship. Positive correlation between episodes means that if the first episode has high intensity, the following episode will have high intensity too. Having negative correlation means that if one episode has high intensity, the intensity of the following episode will diminish. If there is a very weak or no correlation, $r \approx 0$, we say that the intensity of each episode does not affect the intensity of the following episodes.

To compare the two sets of correlations, plots of their 95% credible intervals corresponding to each patient were produced and split into two figures for presentation purposes. Figure 4.1 summarises the correlations for 31 subjects and the remaining subjects are presented in Figure 4.2. The presented intervals are equal tailed credible intervals. Looking at these plots where credible intervals for correlation A, r_A, and correlation B, r_B, are presented by black and red lines respectively, it is clear that most correlations have intervals including 0 meaning that the correlation between the episodes is generally weak. More importantly, there is no considerable difference between the posterior distributions of r_A and r_B, i.e. between the correlation of episodes not occurring within 24 hours, and the correlation of episodes occurring within 24 hours. We conclude that there is no significant difference between those two cases in terms of the time gap between episodes. Only patient 2015 shows some difference in correlation and this might be explained by the fact that this patient has a small number of episodes occurring within 24 hours (seven episodes). We also note here that cases where a red credible interval is not present in the graph, correspond to patients not having episodes occurring within 24 hours.

Next, we want to see if there is any difference in the level of intensity between episodes occurring within 24 hours and episodes not within 24 hours. Again, β_{ik} corresponds to the intensity for episode k for each patient i. From our MCMC run, we obtain the

posterior estimates of β_{ik} for all patients in each iteration. Let Case A be the case where no episode is reported within 24 hours from previous episode and Case B is when the episodes occurred within 24 hours from previous episodes. We are interested in comparing the average posterior intensity between these two cases, i.e. the average over all episodes under consideration for each patient.

Figures 4.3-4.4 show caterpillar plots of the 95% credible intervals of the average posterior intensity for episodes in Case A (black bars) and episodes in Case B (red bars). Again, we have to split the intervals into two figures to accommodate all 66 patients. It is apparent from both plots that the intervals of the two cases overlap thus leading us to conclude that there is no considerable difference in intensity either if episodes occur within 24 hours or not.

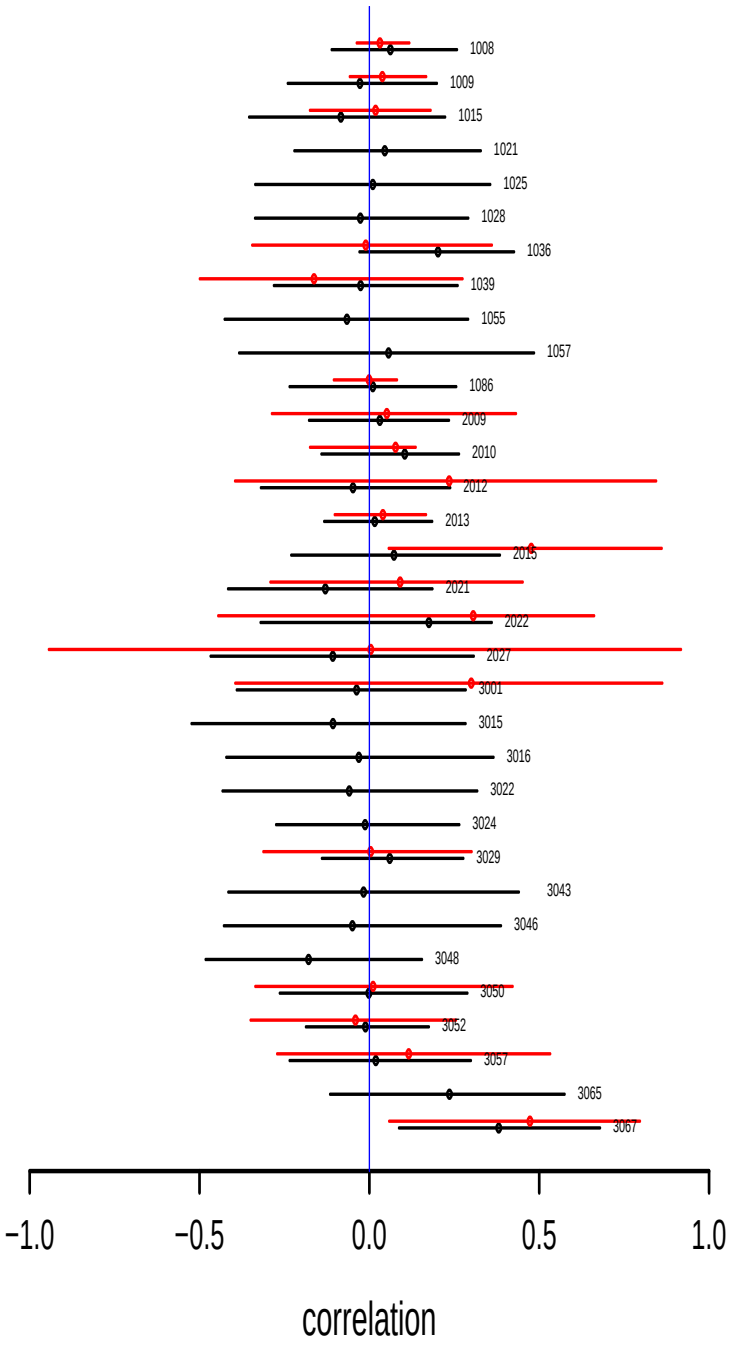


Figure 4.1: Caterpillar plot showing the 95% credible intervals of correlation between the intensity of episodes not within 24 hours (black bars) and between the intensity of episodes within 24 hours (red bars) for patients 1008-3067.

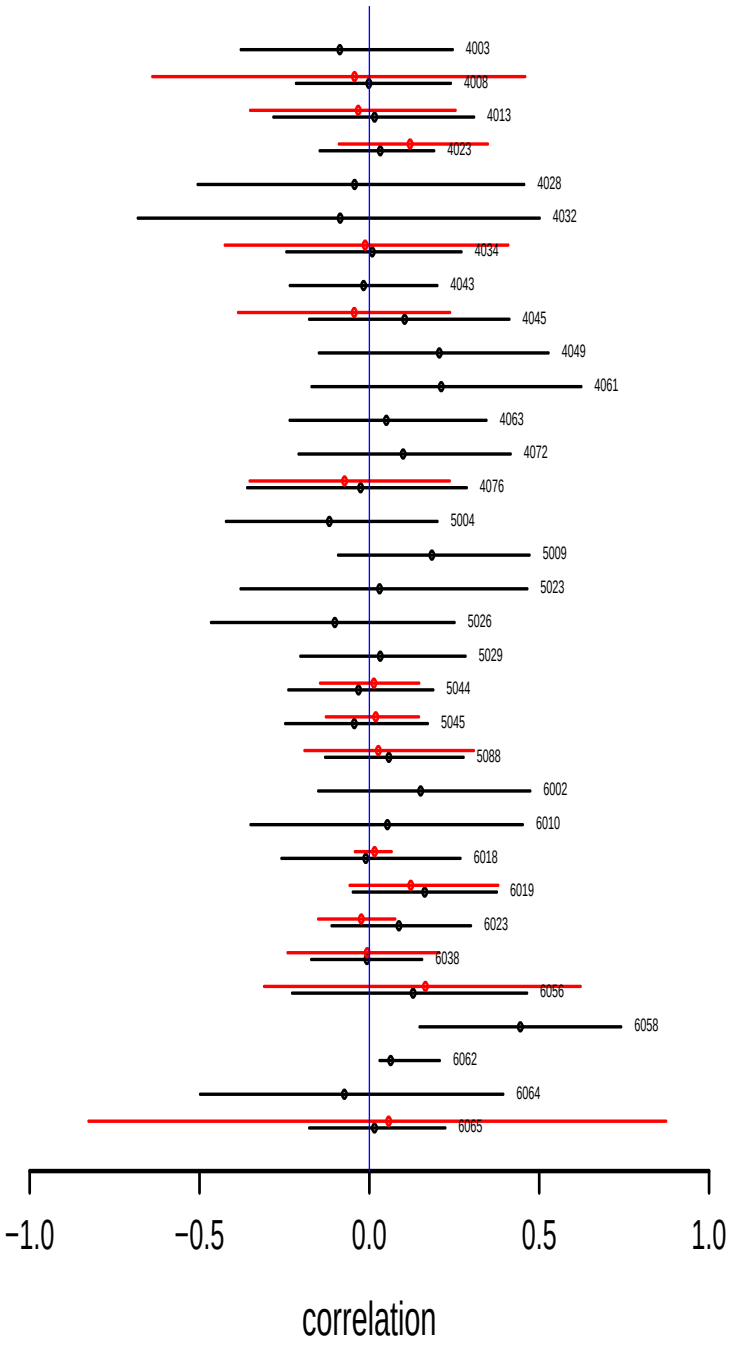


Figure 4.2: Caterpillar plot showing the 95% credible intervals of correlation between the intensity of episodes not within 24 hours (black bars) and between the intensity of episodes within 24 hours (red bars) for patients 4003-6065.

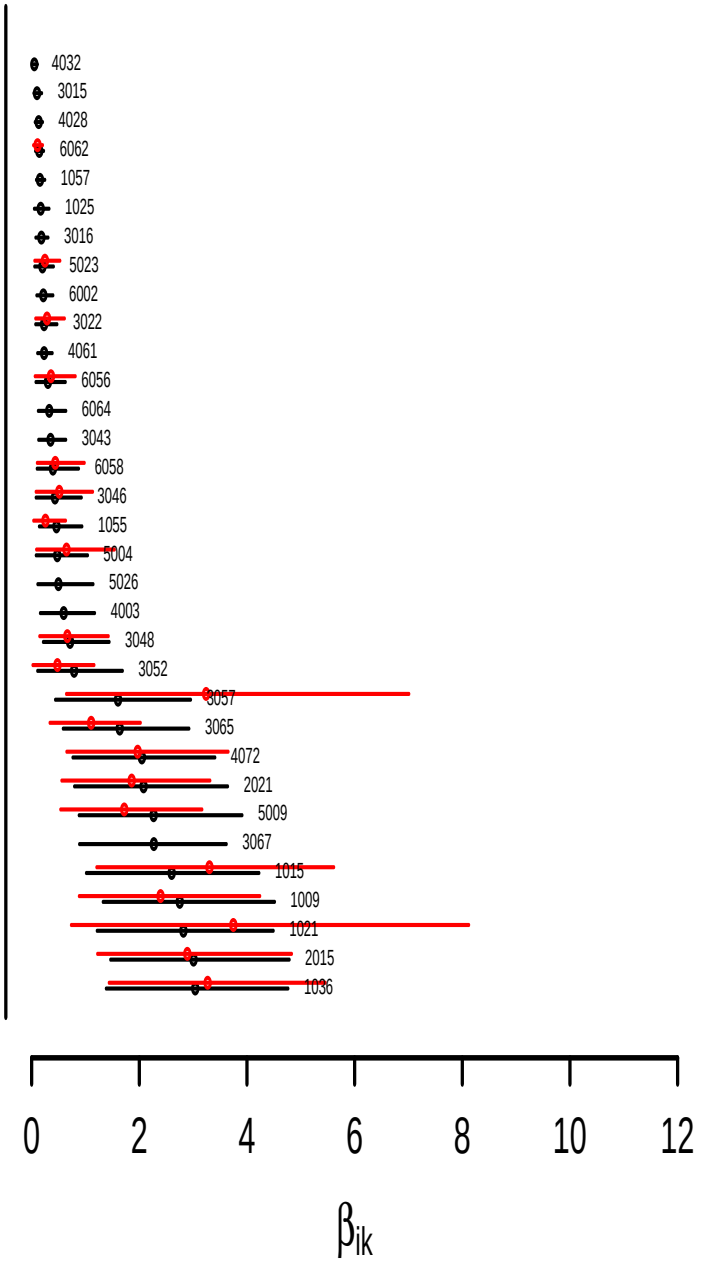


Figure 4.3: Caterpillar plot showing the 95% credible intervals of average posterior intensity for episodes not within 24 hours (black) and episodes within 24 hours (red) for 33 patients, arranged in ascending order based on their posterior mean episode intensity.

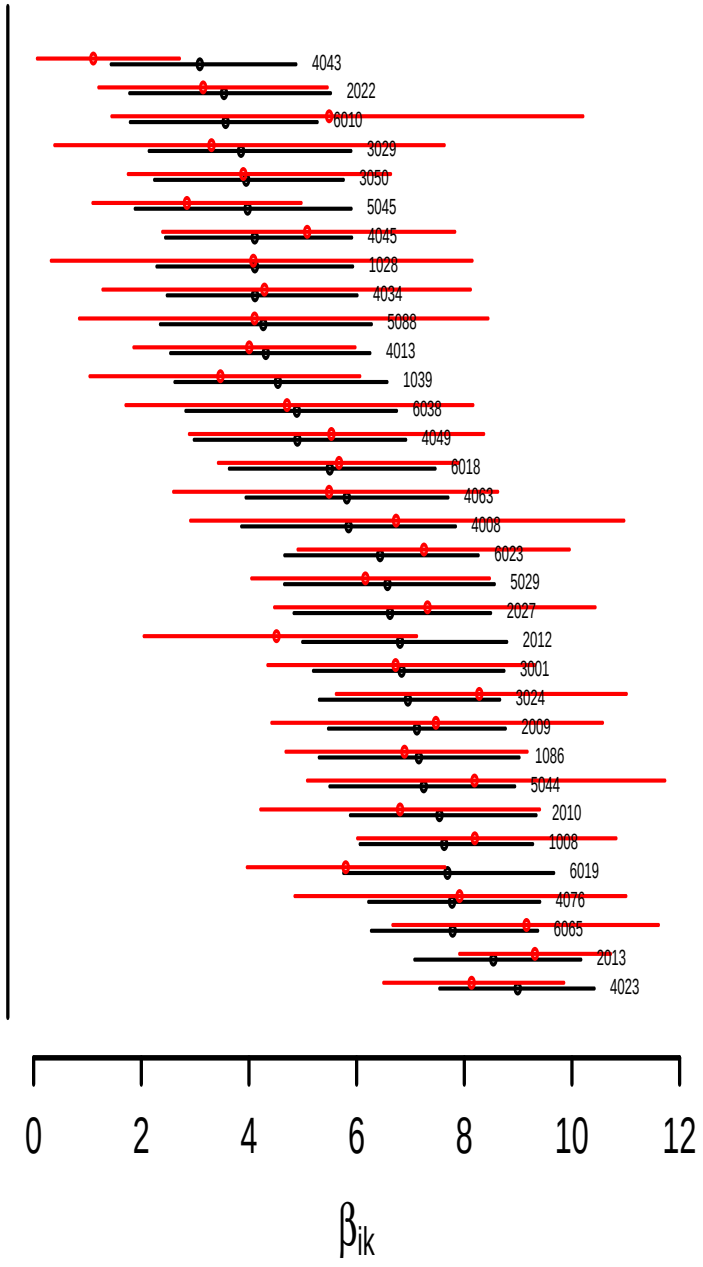


Figure 4.4: Caterpillar plot showing the 95% credible intervals of average posterior intensity for episodes not within 24 hours (black) and episodes within 24 hours (red) for 33 patients, arranged in ascending order based on their posterior mean episode intensity.

4.2 Testing the association between possible episodes

In this section, we want to investigate whether episodes' intensity exhibits change that goes beyond random variation when episodes are in temporal proximity. If there is association between cases where there are no episodes occurring within 24 hours and cases when there are episodes occurring within 24 hours, it is possible that episodes occurring soon after a previous episodes to generate inconsistent symptoms. As previously, β_{ik} is the latent variable that corresponds to the intensity of episode k for patient i. The idea is that if we arrange the posterior estimates of all β_{ik} according to the date and time of the occurrence of the corresponding episode, we should be able to calculate a total 'intensity distance'; from episode 1 to episode K. If we then randomly permute the arrangement of the episodes and calculate the distance again, if there exists association between the intensity of episodes within 24 hours and episodes not within 24 hours, the total intensity distance will become longer. This is because with association, episodes that occurred within 24 hours from a preceding episode may be considered to have big impact on the intensity; either reducing or increasing it.

To implement this idea, we make use of permutation testing. Permutation techniques were introduced by Edgington, (1987) and Blair and Karniski, (1994). Earlier practice of permutation test was given by Fisher in 1971. The study was about the difference between two means of height of different plants (Fisher, 1971).

We are testing the null hypothesis that there is no temporal association in the magnitude of β_{ik} for $k = 1, ..., K_i$ (i.e. the change in β_{ik} over successive episodes is random), while the alternative is that there is such temporal association (i.e. that change in β_{ik} is greater than random change). We run the consistency model with grouped symptoms for 2000 MCMC iterations with the first 1000 iterations discarded as the burn-in. Note that the MCMC convergence is already achieved after the burn-in period. We denote the intensity distance calculated from the posterior β_{ik} estimates as the observed distance, $D^{(obs)}$, whereas the distance calculated using the permuted β_{ik} is referred to as the replicated distance, $D^{(rep)}$.

Figure 4.5 illustrates how we calculate the intensity distance. This plot shows the ordered episodes with their corresponding posterior intensity estimate, $\tilde{\beta}_{ik}$ for patient 4063. The episodes are plotted based on the time gap between them (in days). Black dots and lines refer to $\beta_{4063,k}$ before any permutation (Figure 4.5(a)), while the green colour refers to $\beta_{4063,k}$ after permutation (Figure 4.5(b)). Episodes occurring within 24 hours from a previous episode are plotted red. Consider the intensity distance, d_{ik}, between two successive episodes, $k-1$ and k, for patient i, with corresponding intensities $\beta_{i,k-1}$ and $\beta_{i,k}$. If we also assume that the two episodes occur at times $t_{i,k-1}$ and $t_{i,k}$ respectively, then the intensity distance between them is

$$d_{ik} = \sqrt{(t_{i,k} - t_{i,k-1})^2 + (\tilde{\beta}_{i,k} - \tilde{\beta}_{i,k-1})^2}. \tag{4.2.1}$$

The total intensity distance for patient i is given as

$$D_i = \sum_k d_{ik} \text{ , where } d_{ik} \text{ is (4.2.1).}$$

The algorithm is summarised as follows.

For patient i, at each $\gamma = 1,...,1000$ iteration,

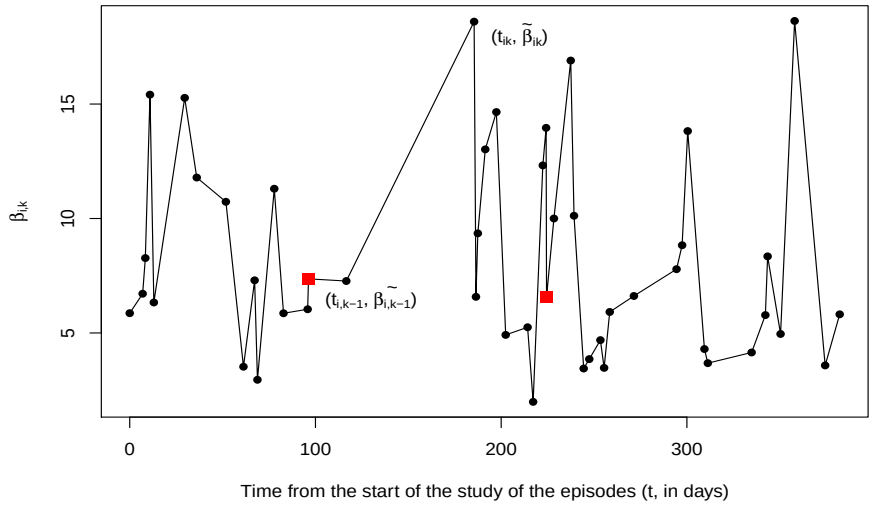
i. Compute the observed distance, $D_{i,\gamma}^{(obs)}$.

ii. Permute the $\tilde{\beta}_{ik}$'s n times. Calculate the new replicated distance, $D_i^{(rep)}$ each time.

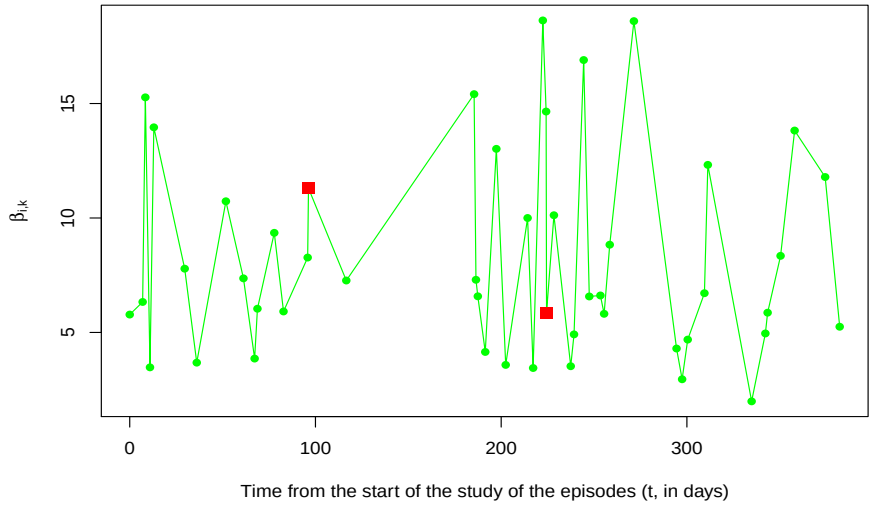
Step ii is repeated for $n = 100$ iterations, thus using the permutation of β_{ik} will generate a distribution of the test statistic under the null hypothesis (Edgington, 1987 and Manly, 1997). From each sampling distribution, we obtain a p-value, π_γ. Under the alternative hypothesis, we expect large D_i values, hence

$$\pi_\gamma = \frac{\text{number of } \left\{ D_i^{(rep)} > D_{i,\gamma}^{(obs)} \right\}}{\text{number of } D_i^{(rep)}}.$$

Therefore, after 1000 iterations we can produce a histogram of p-values, π_γ, for every patient and observe any evidence against the tested hypothesis. We present histograms for six subjects selected at random (Figure 4.6). Observing the distribution



(a) Before permutation

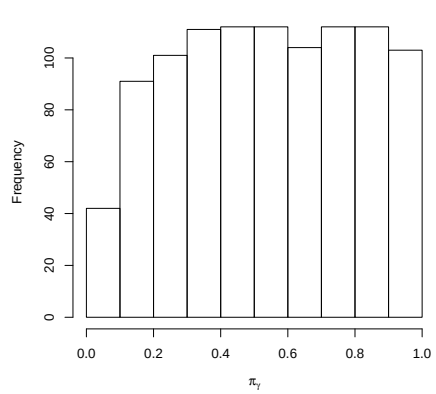


(b) After permutation

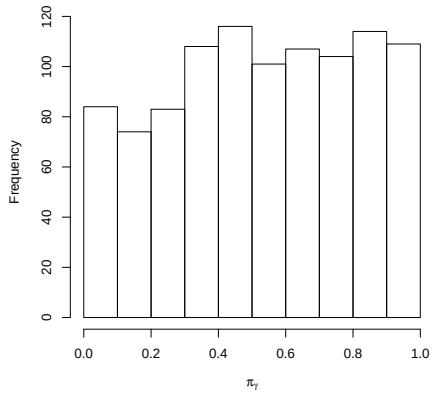
Figure 4.5: This graph plots the posterior estimates of intensity versus the time gap between episodes (in days) for patient 4063. Red points refer to the episodes occurring within 24 hours from previous episode. Black plots and lines refer to $\beta_{4063,k}$ before any permutation and the green colour refers to $\beta_{4063,k}$ after permutation.

of π_γ, patient intensities do not show any evidence against the tested hypothesis. The null hypothesis will be rejected if we have very small π_γ, for example smaller than 0.05. Therefore, we are interested to see the proportion of π_γ less than 0.05, $\Pr(\pi_\gamma < 0.05)$. Figure 4.7 plots the $\Pr(\pi_\gamma < 0.05)$ versus the mean time between episodes for each patient. Patients with many episodes occurring within 24 hours are likely to have smaller mean time between their episodes. However, there is no specific pattern observed from this plot. There are four subjects with high $\Pr(\pi_\gamma < 0.05)$; 1057, 1008, 4045, and 5045. Histograms of p-values for these four patients are shown in Figure 4.8. Evidence against the null hypothesis here is stronger than in Figure 4.6, as the plots exhibit higher proportion of $\Pr(\pi_\gamma < 0.05)$. Especially, patient 1057 has the highest $\Pr(\pi_\gamma < 0.05) = 0.796$. However, this subject has no episodes within 24 hours and average mean time between episodes = 25.17. For patient 5045 the ratio of episodes within 24 hours over all episodes is 0.13 and the average mean time between episodes is 4.03 days. $\Pr(\pi_\gamma < 0.05)$ for this patient is 0.243. Subject 4045 has 1 episode occurring within 24 hours from a total of 36 episodes. The mean time between episodes is 17.53 days with $\Pr(\pi_\gamma < 0.05) = 0.3180$. Subject 1008 has mean time between episodes of 3.23 days and $\Pr(\pi_\gamma < 0.05) = 0.408$. Patient 1057 has the highest $\Pr(\pi_\gamma < 0.05)$: 0.7960. This subject has no episodes within 24 hours and average mean time between episodes = 25.17.

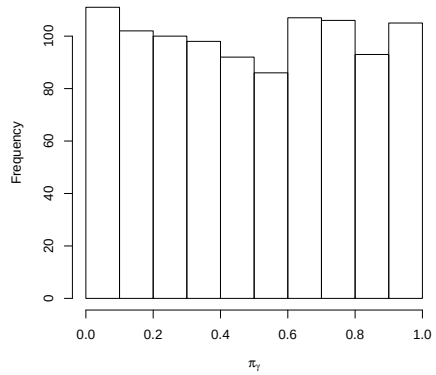
As a conclusion, we can argue that the intensity of episodes that occur within 24 hours is not systematically different from the intensity of other episodes, and does not exhibit higher correlation either. Therefore, it is reasonable to include those discarded episodes into our analysis.



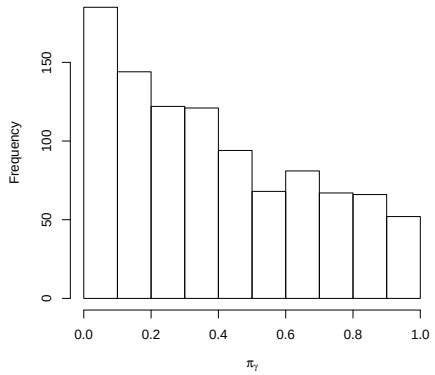
(a) 2012



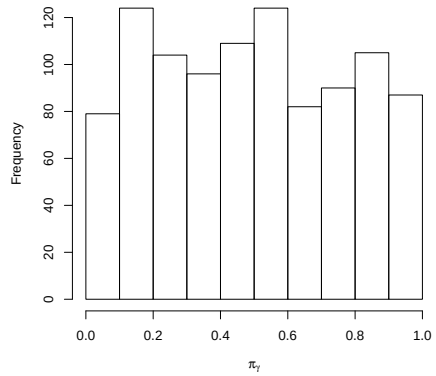
(b) 2013



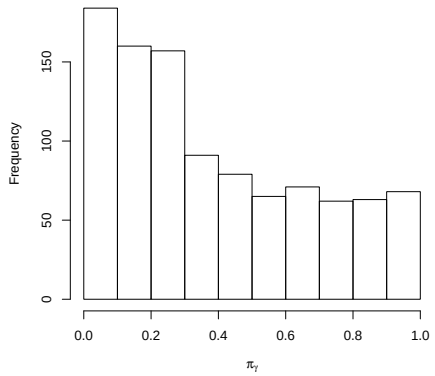
(c) 2021



(d) 3057



(e) 5009



(f) 3043

Figure 4.6: Histograms of p-values, π_γ, when testing the association between the intensity of episodes for Subjects 2012, 2013, 2021, 3057, 5009, and 3043.

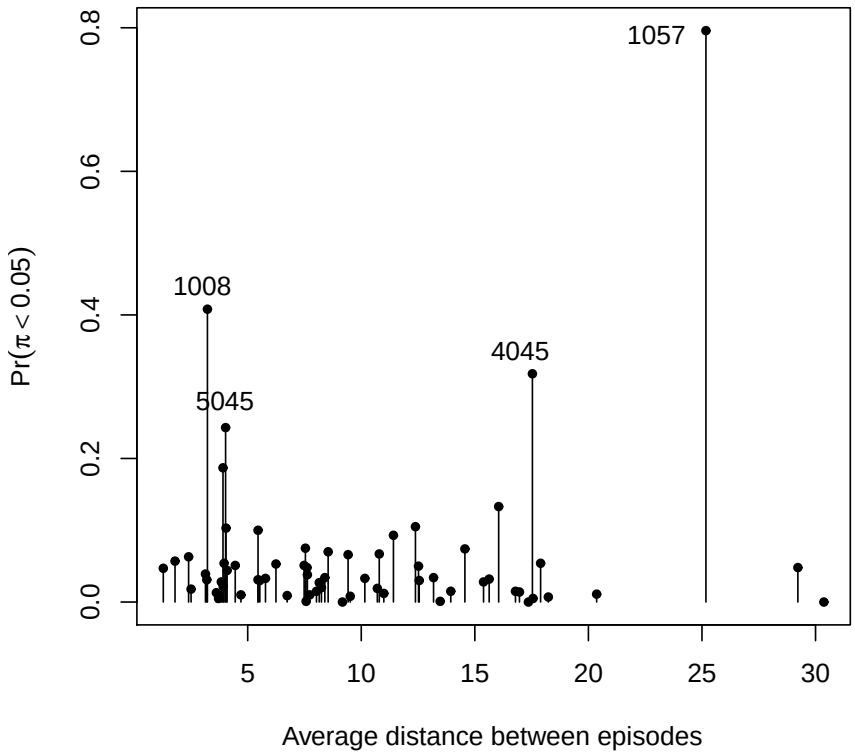
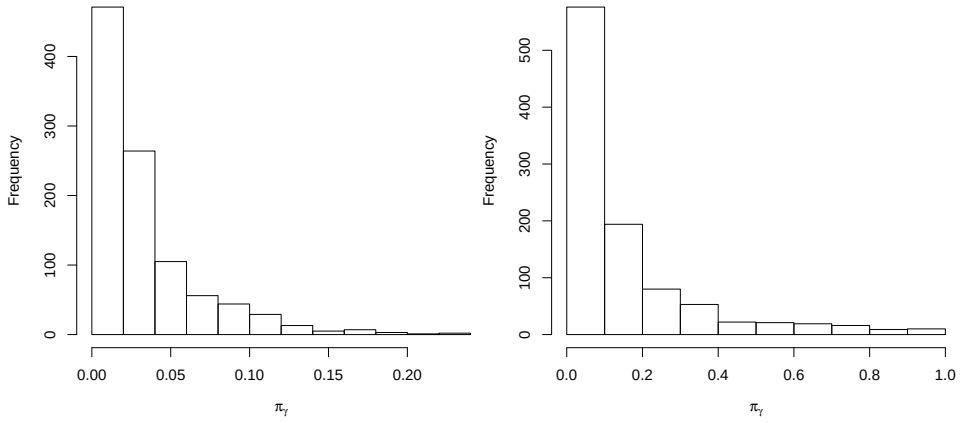
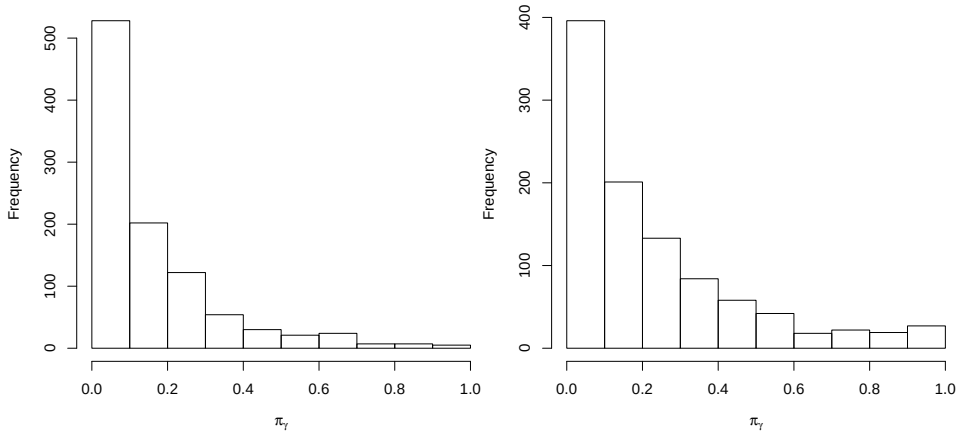


Figure 4.7: Plot of the Pr(π_γ <0.05) versus the mean time between episodes for each patient.



(a) 1057 (b) 1008



(c) 4043 (d) 5045

Figure 4.8: Histograms of p-values, π_γ, when testing the association between the intensity of episodes for Subjects 1057, 1008, 4043 and 5045.

4.3 Considering episodes within 24 hours in the analysis

In previous sections, we investigated episodes reported within 24 hours from a previous episode and concluded that these episodes do not result in diminished intensity of hypoglycaemia. Thus, there is no reason to exclude those episodes from further analysis. Hence, episodes with missing date of occurrence can also be included in our work as there is no need to determine the time gap between consecutive episodes. This section will discuss how these episodes affect consistency when they are included in our work.

Eleven subjects do not have any episode occurring within 24 hours or episode with missing date. Therefore, this section will focus on the remaining 55 subjects. Table 4.1 summarises the number of episodes for each patient together with their number of episodes that occurred within 24 hours and episodes with missing date. Three different measures or estimates are used to show the effect of adding those episodes: relative difference in consistency; the coefficient of variation of consistency; and consistency estimates. All results and discussions are based on the model with grouped symptoms.

Relative difference in consistency

We use the relative difference in consistency estimates after episodes within 24 hours and episodes with missing date are included in our work to assess possible changes. Let RDC_i denote the relative difference in consistency for patient i, given as

$$RDC_i = \frac{\tilde{c_i}^{all} - \tilde{c_i}}{\tilde{c_i}^{all}},$$

where $\tilde{c_i}^{all}$ denotes the posterior estimate of consistency for patient i when all episodes are included in the analysis, and $\tilde{c_i}$ is the corresponding estimate when episodes occurring within 24 hours or episodes with missing date are excluded. Figure 4.9 plots the relative difference of estimated consistency when those episodes are included in the analysis versus the corresponding percentage of those episodes for each subject.

Generally, the majority of the patients have relative mean difference in consistency about zero, indicating little change in their consistency estimates. Only six subjects have quite a distinct reduction in the consistency estimate. These are subjects 2013 (55% within 24 hours episodes), 2015 (32%), 5004 (14%), 5044 (21%), 6023 (22%) and 6038 (37%). Negative values in relative difference of consistency suggest that these patients' consistency was lower after adding the within 24 hours episodes. While other patients also have high percentage of those episodes, only these six patients experience big decrement in consistency.

Patient 2015 has relative difference in consistency of -1.1853. Since relative difference is usually expressed in percentage, we say that for this patient the consistency estimated from using all episodes is 118% lower than the consistency estimated when we exclude the within 24 hours episodes and episodes with missing date. Subject 2013 has relative difference in consistency of 107% while the relative difference for subjects 5004, 5044, 6023, and 6038 are 145%, 92%, 171%, and 127% respectively.

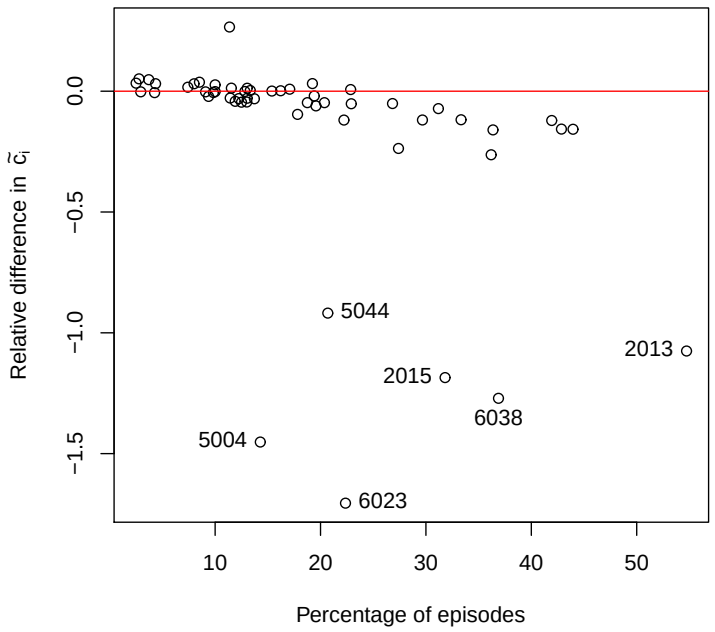


Figure 4.9: Relative difference in consistency estimates versus the percentage of episodes occurring within 24 hours and episodes with missing date.

Coefficient of variation

We also consider the consistency's coefficient of variation (CV) for each patient in both situations; with and without the episodes occurring within 24 hours from previous episode and episodes with missing date. The CV is defined as:

$$CV = \frac{\text{SD}(\tilde{c}_i)}{\text{mean}(\tilde{c}_i)},$$

where $\tilde{c}_i$ is the posterior consistency estimate of patient i. Our interest is to see in which case the consistency is more dispersed. This is measured by calculating the difference between the CV in two cases;

$$CV_i^{all} - CV_i,$$

where CV_i^{all} is the CV for patient i when all episodes are included and CV_i is when episodes within 24 hours or episodes with missing date are excluded. The difference of these CVs for every patient is plotted against their percentage of episodes within 24 hours and episodes with missing date in Figure 4.10. Negative CV difference implies that the calculated CV when including all episodes is smaller than CV without those episodes. The majority of subjects we observed had negative CV difference implying that the calculated CV when including those episodes is smaller than the CV without those episodes (with the difference being rather small, i.e. ≤ 0.02). This could also be interpreted as more reliable estimates (smaller CV) when the additional episodes are included in the analysis.

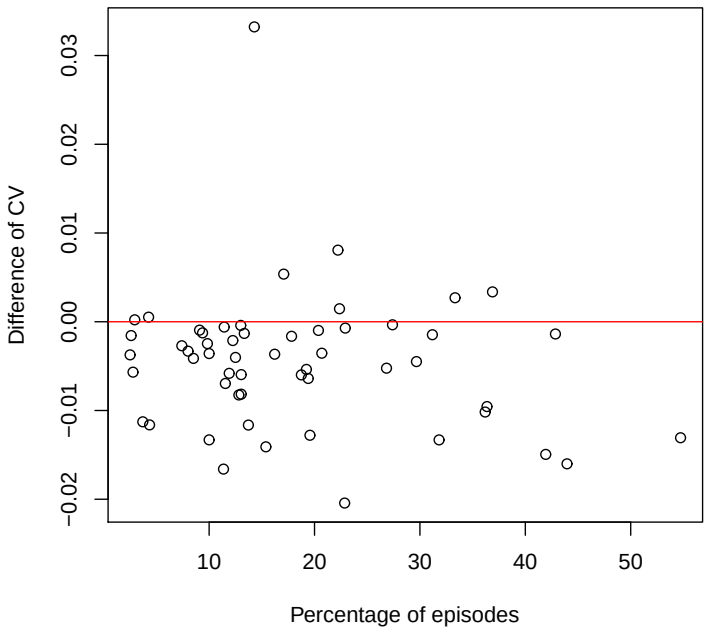


Figure 4.10: Difference in consistency's coefficient of variation versus the percentage of episodes occurring within 24 hours.

Subject	Total number of all episodes	Number of episodes within 24 hours	Number of episodes with missing date	% within 24 hours and missing date episodes
1008	123	25	8	26.83
1009	96	19	3	22.92
1015	35	8	0	22.86
1021	49	1	5	12.24
1025	26	0	5	19.23
1028	45	2	4	13.33
1036	42	5	13	42.86
1039	47	3	1	8.51
1055	27	2	0	7.41
1057	11	0	0	0
1086	91	24	3	29.67
2009	101	10	8	17.82
2010	102	13	1	13.73
2012	42	5	0	11.9
2013	296	140	22	54.73
2015	44	7	7	31.82
2021	33	6	5	33.33
2022	41	5	2	17.07
2027	26	3	1	15.38
3001	35	4	0	11.43
3015	19	0	0	0
3016	23	3	0	13.04
3022	23	2	1	13.04
3024	30	1	2	10
3029	78	7	2	11.54
3043	17	0	0	0
3046	25	2	0	8
3048	37	0	6	16.22
3050	54	8	3	20.37
3052	146	29	11	27.4
3057	91	21	19	43.96
3065	27	1	0	3.7
3067	32	6	0	18.75
4003	26	0	0	0
4008	67	5	8	19.4
4013	55	11	9	36.36
4023	210	64	12	36.19
4028	12	0	0	0
4032	10	0	0	0
4034	61	6	0	9.84
4043	36	1	0	2.78
4045	46	6	3	19.57
4049	23	1	0	4.35
4061	12	0	0	0
4063	48	2	4	12.5
4072	40	1	0	2.5
4076	32	3	0	9.38
5004	28	1	3	14.29
5009	40	1	3	10
5023	22	2	0	9.09
5026	26	0	0	0
5029	47	2	0	4.26
5044	87	14	4	20.69
5045	78	10	0	12.82
5088	100	11	2	13
6002	34	1	0	2.94
6010	19	0	0	0
6018	62	21	5	41.94
6019	93	27	2	31.18
6023	76	17	0	22.37
6038	122	40	4	36.89
6056	36	6	2	22.22
6058	20	0	0	0
6062	13	1	0	7.69
6064	17	0	0	0
6065	44	3	2	11.36

Table 4.1: Number of hypoglycaemic episodes for individual patients.

Consistency estimates

To compare the distributions of $\tilde{c}_i$ between the two cases, we plot histograms of these cases on the same plot in Figure 4.11. The overlapping histograms tell us that the distribution of $\tilde{c}_i$ does not significantly change after we include the 24 hour episodes and/or episodes with missing date. The mean of $\tilde{c}_i$ when we exclude the episodes within 24 hours or episodes with missing date (yellow plot) is 43.39 with standard deviation of 18.41 whereas the mean of $\tilde{c}_i$ when using all episodes (blue plot) is 45.26 with standard deviation of 18.74.

In addition, when comparing the boxplots of the two cases, we observe that the two plots are very similar in shape and scale (Figure 4.12). The red boxplot represents the estimated consistency when using all episodes and the quartiles of $\tilde{c}_i$ are $q_0 = 7.82$, $q_{0.25} = 30.55$, $q_{0.5} = 43.82$, $q_{0.75} = 57.03$ and $q_1 = 95.62$. When excluding the episodes occurring within 24 hours or episodes with missing date, the quartiles of $\tilde{c}_i$ obtained are $q_0 = 7.20$, $q_{0.25} = 30.76$, $q_{0.5} = 43.30$, $q_{0.75} = 55.90$ and $q_1 = 95.54$ (yellow plot). Table 4.2 provides the summary of posterior estimates of $\tilde{c}_i$ for each subject when including all episodes. The summary of the corresponding estimates when we exclude episodes within 24 hours and no date is already given in Table 3.7.

We then produce violin plots to compare the estimated posterior distributions of consistency for each patient in both cases (with and without episodes occurring within 24 hours and episodes with missing date) in Figures 4.13-4.14. These plots are divided into two plots such that Figures 4.13 contains the first 27 subjects listed in ascending order of their unique code number, while the plots for the remaining subjects are given in Figure 4.14. Red plots refer to the estimates obtained when we include the within 24 hours episodes and episodes with missing date in the model whereas yellow plots refer to the estimates when those episodes were excluded. Examining these plots, we observe that overall there is no significant difference in the distribution of consistency between the two cases. The majority of the patients have very similar densities in the two cases. However, there are six subjects with some difference in their posterior distribution of consistency. These are the same patients as identified previously (when looking at relative difference in consistency), i.e. patients 2013, 2015, 5004,

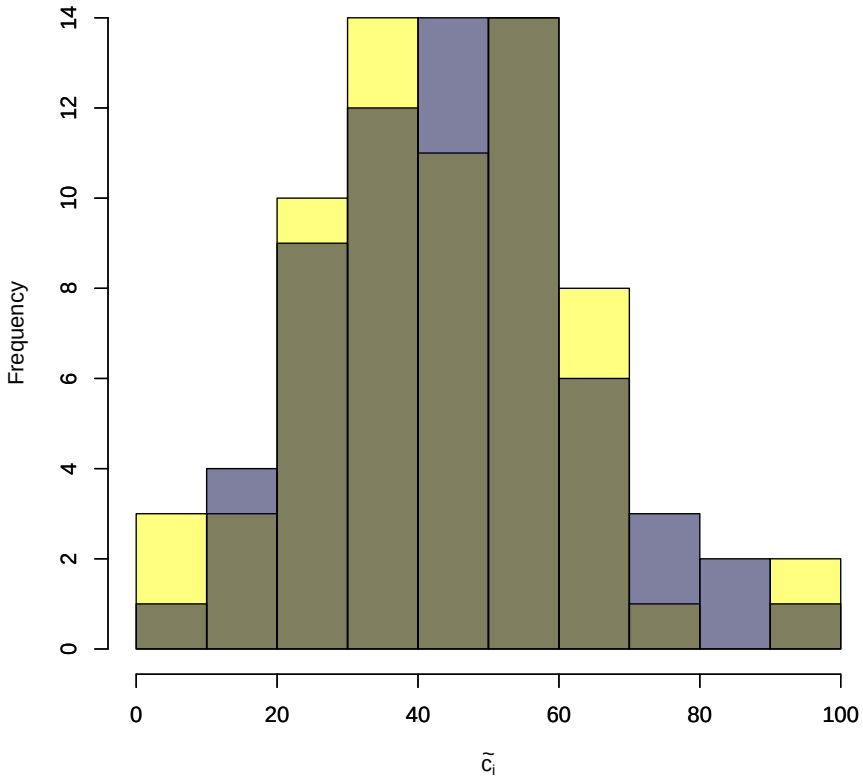


Figure 4.11: Histograms of posterior consistency. The grouped symptoms model is used with all episodes (blue) and without the 24h episodes or episodes with missing data (yellow). Grey colour is used to highlight the overlap region between the two distributions.

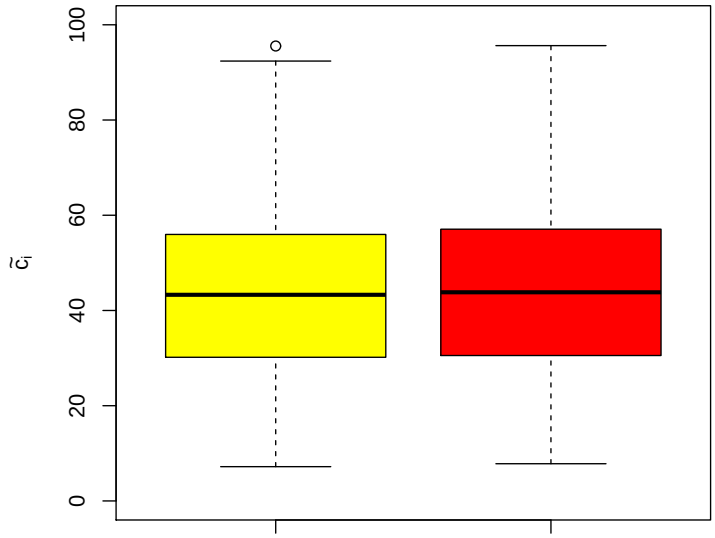


Figure 4.12: Boxplots of posterior consistency. The grouped symptoms model is used with all episodes (red) and without the 24h episodes or episodes with missing data (yellow).

5044, 6023, and 6038. The differences in their posterior estimate of $\tilde{c}_i$ are listed in Table 4.3.

As given before, subject 2013 has 54.33% new episodes of the total number of episodes. When using all available episodes, the posterior consistency density shifts to the left and has narrower distribution, as compared to when excluding those new episodes (Figures 4.13). As for subject 2015, from all 44 episodes 31.82% of them are newly added. Again, the scales of the density when using all episodes is much narrower and also moved to the left (Figure 4.13).

With subjects 5004, 5044, and 6023 we observe the same trend, where their consistency densities are much narrower and shifted to the left when the new episodes are considered (Figure 4.14). Another subject with similar result is subject 6038. This patient has total number of episodes of 122 and 31.97% of them are new episodes. Although the distributions in the two cases are different for these six subjects, the densities overlap indicating that there is no significant difference in consistency be-

tween these two cases.

Subject	$\tilde{c}_i$	SD	25% CI	95% CI
1008	26.19	3.244	19.99	32.77
1009	50.28	4.817	41.06	60.17
1015	32.99	5.682	22.60	45.00
1021	43.08	5.901	32.41	54.87
1025	43.52	7.703	29.64	59.66
1028	95.62	2.194	89.93	98.39
1036	51.32	6.408	38.95	63.87
1039	32.66	5.525	22.98	44.25
1055	56.91	7.811	41.55	72.44
1057	83.72	6.158	69.47	93.00
1086	27.97	3.745	21.18	36.01
2009	23.95	3.250	18.00	30.74
2010	25.76	3.511	19.36	33.17
2012	27.46	4.797	19.03	37.80
2013	7.78	0.890	6.13	9.63
2015	15.81	2.460	11.48	21.25
2021	42.80	6.651	30.47	56.24
2022	37.92	5.538	27.70	49.44
2027	49.22	7.855	34.56	64.89
3001	47.41	6.924	34.45	61.76
3015	15.93	3.052	10.77	22.59
3016	62.43	7.598	46.92	76.85
3022	50.60	7.959	35.66	66.65
3024	58.22	7.006	44.86	72.04
3029	74.79	4.692	64.67	83.29
3043	42.45	8.583	26.05	59.52
3046	50.77	7.640	36.48	66.04
3048	60.00	6.715	46.80	73.27
3050	53.30	5.540	42.78	64.40
3052	41.41	3.939	33.99	49.25
3057	29.55	3.920	22.39	37.61
3065	38.65	6.884	26.14	53.03
3067	45.88	6.498	33.51	58.89
4003	62.67	6.929	49.05	75.70
4008	49.20	4.849	39.76	58.67
4013	30.54	4.523	22.27	39.92
4023	20.74	2.247	16.56	25.41
4028	53.86	9.885	34.67	72.66
4032	75.70	9.474	54.65	90.85
4034	64.89	5.198	54.54	74.48
4043	30.11	5.048	20.91	40.83
4045	33.79	5.473	23.67	45.14
4049	57.55	7.579	42.61	72.02
4061	78.34	6.810	63.50	89.60
4063	27.55	4.516	19.54	37.09
4072	44.01	6.536	31.80	57.17
4076	39.94	6.563	28.12	53.64
5004	18.16	3.630	12.08	26.24
5009	48.81	6.168	37.25	61.08
5023	51.23	8.528	35.42	68.55
5026	64.78	7.179	50.26	78.40
5029	60.83	6.112	48.58	72.60
5044	16.80	2.170	12.94	21.65
5045	42.21	4.978	32.79	52.30
5088	18.74	2.769	13.72	24.74
6002	57.07	6.291	44.75	69.45
6010	55.33	8.347	38.87	71.27
6018	37.48	4.631	28.86	47.13
6019	30.10	3.952	22.74	38.33
6023	10.41	1.640	7.51	13.95
6038	12.47	1.740	9.33	16.06
6056	58.28	6.469	45.20	70.52
6058	69.58	7.362	54.09	82.86
6062	49.37	9.668	30.74	68.66
6064	88.90	4.156	79.24	95.22
6065	23.76	5.520	14.68	36.29

Table 4.2: Posterior estimates of consistency (Mean, standard deviation, and 95% credible interval) for all subjects using grouped symptoms model when including all episodes.

	without 24h & no date episodes			all episodes		
Subject	mean	SD	95% CI	mean	SD	95% CI
2013	16.15	2.06	(12.43, 20.59)	7.78	0.89	(6.13, 9.63)
2015	34.55	5.83	(24.24, 46.73)	15.81	2.46	(11.48, 21.25)
5004	44.53	7.42	(31.47, 58.51)	18.16	3.63	(12.08, 26.24)
5044	32.23	4.29	(24.20, 41.11)	16.80	2.17	(12.94, 21.65)
6023	28.16	4.39	(20.46, 37.56)	10.41	1.64	(7.51, 13.95)
6038	28.32	3.85	(21.64, 37.15)	12.47	1.74	(9.33, 16.06)

Table 4.3: Posterior estimates of $\tilde{c}_i$ when using the grouped symptoms model with all episodes and without episodes occurring within 24 hours of preceding episode and episodes with missing date.

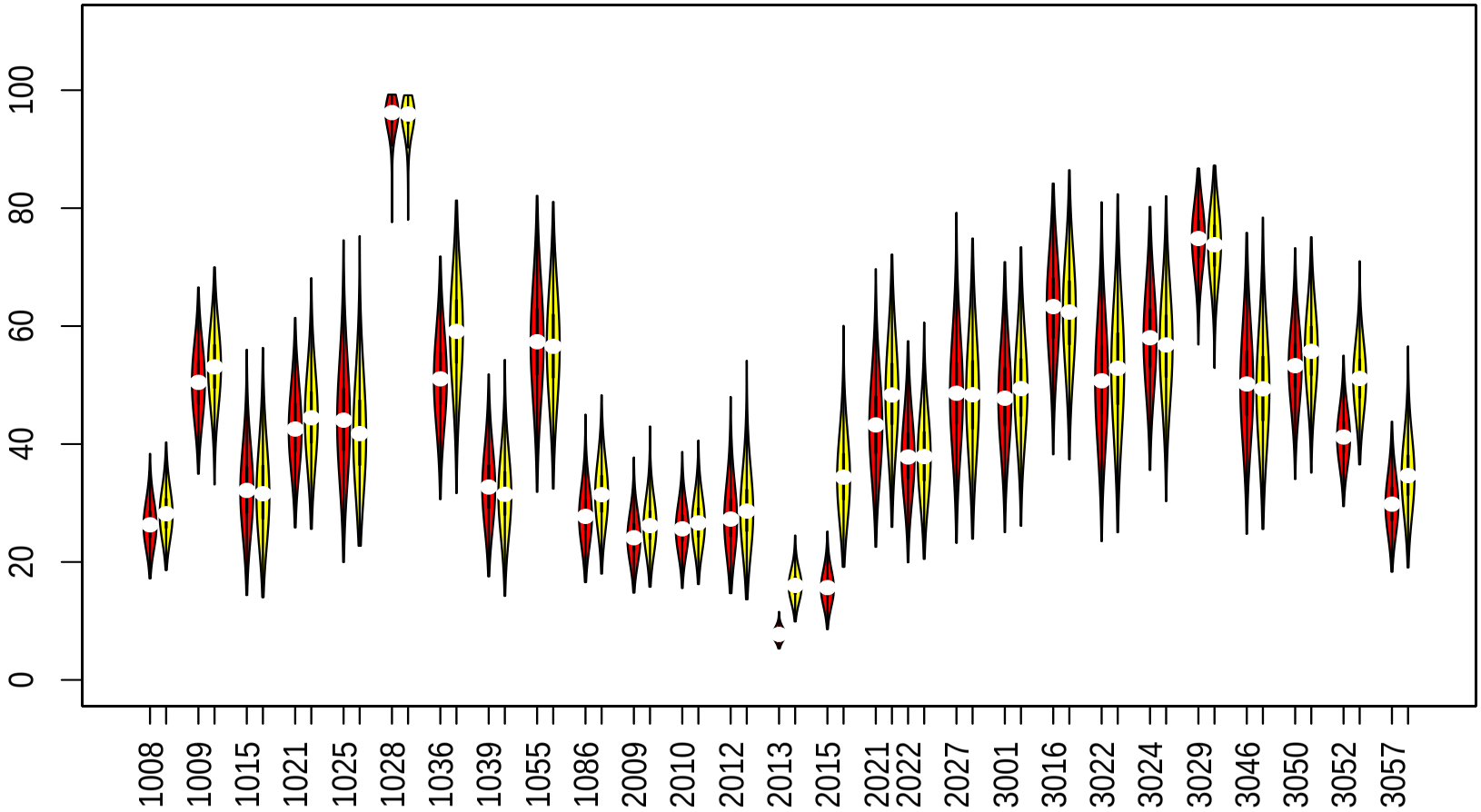


Figure 4.13: Violin plots of posterior consistency. The grouped symptoms model is used with (red) and without (yellow) episodes occurring within 24h of preceding episodes and episodes with missing date.

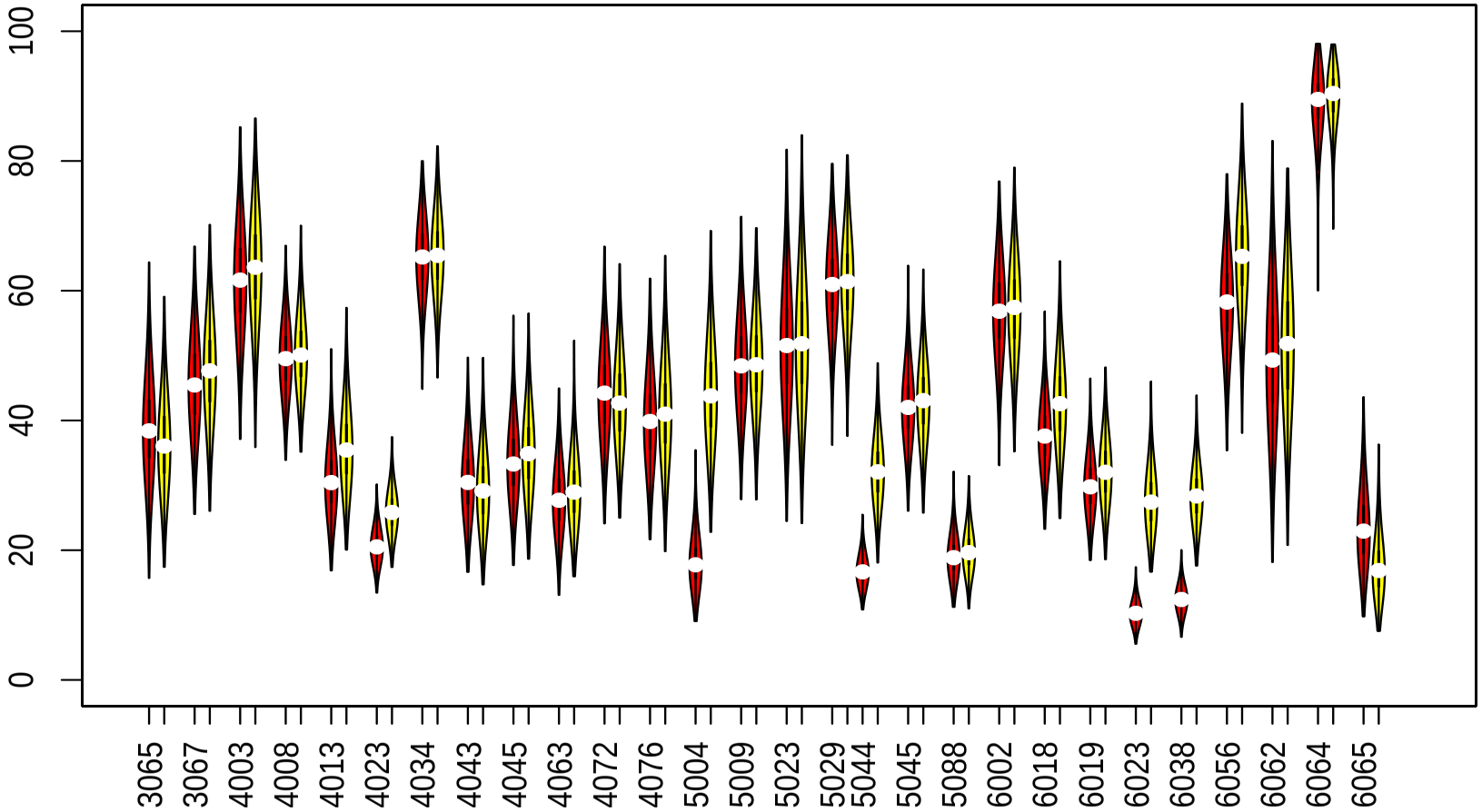


Figure 4.14: Violin plots of posterior consistency. The grouped symptoms model is used with (red) and without (yellow) episodes occurring within 24h of preceding episodes and episodes with missing date.

4.4 Concluding remarks

Previous work had excluded episodes in temporal proximity from further analysis because there is a possibility that these episodes may significantly affect the intensity of other episodes. This chapter explores different ways to assess the effect of including these episodes in our study. First by assessing the correlation of the intensity of episodes and difference in posterior intensity of the episodes between the two cases; including and excluding episodes within 24 hours does not support the claim that there are any significant changes in intensity. Furthermore, the permutation test conducted also concludes that the change in the posterior estimates of β_{ik} over successive episodes is random, i.e. there is no significant difference in intensity between the two cases. By using all episodes in the analysis, we also discussed the relative difference in consistency, the coefficient of variation and the consistency estimates between the two cases; with and without the within 24 hours episodes. Therefore, our analysis suggests that episodes occurring within 24 hours from their previous episode are generally not significantly affected by the intensity of preceding hypoglycaemic episodes, and in all but six cases do not result in a considerable difference in individual patients' consistency.

Chapter 5

Association Between Consistency and Patient-Specific Covariates

In this chapter, we no longer exclude episodes within 24 hours from our analysis. Based on the conclusion we made in previous chapters, episodes that occur soon after other episodes do not affect considerably the intensity of the episodes of hypoglycaemia. Thus, the analysis in this chapter uses the complete data set of reporting regardless of the occurring time. This chapter discusses the investigation of what factors affect the consistency of symptom reporting of individual patients. There are ten factors in consideration and we also test the effect of interactions between them. The association between the factors and consistency is modelled with a Bayesian Generalised Linear Model (GLM), using the individuals' posterior precision estimates obtained from the model with grouped symptoms discussed in Section 4.3 as the response variable. Also, in this chapter, we expand the consistency estimation model of Section 3.4 to allow hierarchical modelling where we are able to do both estimations in one model: estimate consistency and identify significant factors affecting consistency.

5.1 Modelling using estimates from grouped symptoms model

To help diabetic patients recognise their hypoglycaemic episodes efficiently, it is vital to educate them about factors that may cause variation in their symptoms. Zammit et al. (2011) previously investigated the effect of ten patient-specific covariates on

consistency using GLM methodology. Here we expand this analysis by first including episodes occurring within 24 hours from a previous episode and episodes with missing date. The covariates we use are gender, age, diabetes type (1 or 2), duration of diabetes, retinopathy, hypoglycaemia awareness score (1 to 7, with higher scores corresponding to weaker awareness of hypoglycaemia), body mass index (BMI), stimulated C-peptide, haemoglobin A1c (Hba), and serum angiotensin converting enzyme activity (ACE). A brief explanation on each covariate is already provided in Chapter 1 and details of the values of these covariates are summarised in Table 5.1.

	Covariates	Number of Levels	Measurement units/ Additional Information
x_1	Age	Numerical	Years
x_2	Duration of diabetes	Numerical	Months
x_3	BMI	Numerical	kg/m^2
x_4	C-peptide	Numerical	nmol/L
x_5	Hba	Numerical	Percent
x_6	ACE	Numerical	IU/L
x_7	Awareness of hypoglycaemia	7	1=aware , 7=not aware
x_8	Gender	2 (M & F)	M is the base category
x_9	Type of diabetes	2 (T1 & T2)	T1 is the base category
x_{10}	Retinopathy	3	1. No retinopathy 2. Background retinopathy 3. Proliferative retinopathy

Table 5.1: Definitions of the covariates

All covariates have numerical values except for gender, awareness of hypoglycaemia, type of diabetes and retinopathy which are considered as categorical factors in the model. For the categorical factors gender and type of diabetes, each factor has one coefficient in the model with male and Type 1 as the baseline (i.e. gender was coded as male = 0 and female = 1, while Type 1 diabetes = 0 and Type 2 diabetes= 1). As for retinopathy, it has three coefficients and we use a sum-to-zero constraint (i.e. $b_{10,1}RET_1 + b_{10,2}RET_2 + b_{10,3}RET_3 = 0$) for comparing effects to a mean level. RET1 was also coded as binary, 0 when subjects have no retinopathy and 1 otherwise. RET2 was coded as 0 when subjects have background retinopathy and 1 otherwise. For RET3, proliferative retinopathy = 1 and otherwise = 0.

Without standardising, Figure 5.1 shows the box plots of the six covariates with numerical values; age, BMI, duration, C-peptide, Hba, and ACE. The mean of age and

BMI are 54.45 and 26.42 respectively. Data for duration are in months and have mean of 205.64 with the minimum duration is six months. The mean for C-peptide is 0.50 with the highest reading being 2.51. Hba and ACE both have mean of 7.46 and 32.47 respectively.

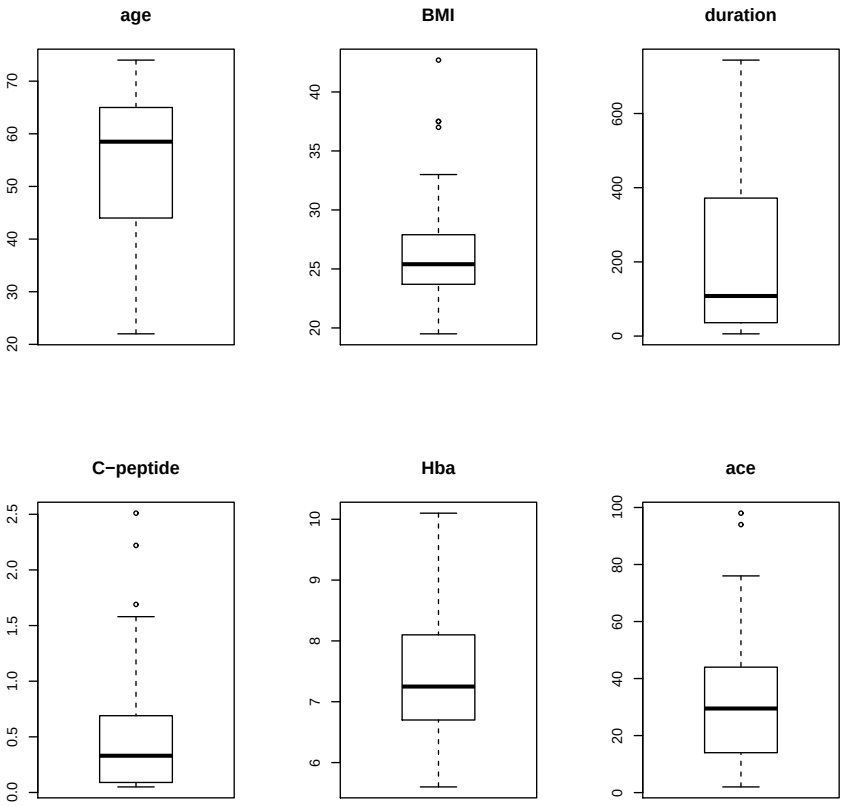


Figure 5.1: Box plots of six covariates with numerical values.

We utilise the posterior precision estimates, $\tilde{\sigma}_i^{-2}$ obtained from the grouped symptoms model in Section 4.3. Based on our discussion in Chapter 4, we include data from all episodes reported regardless of the time they occur. To link estimates of the posterior estimate of the consistency parameter σ_i^{-2} with the covariates, a generalised linear model with gamma errors is used, with linear predictor:

$$\log\{E(\sigma_i^{-2})\} = b_0 + \sum_{l=1}^{7} b_l z_{il} + b_8 GEN_i + b_9 TYPE_i$$

$$+ b_{10,1} RET_1 + b_{10,2} RET_2 + b_{10,3} RET_3 \qquad (5.1.1)$$

for $i = 1, ..., 66$, GEN represents gender, TYPE represents type of diabetes, and RET represents covariate retinopathy. z_{il} are the standardised observations of covariates $x_1, ..., x_7$ in Table 5.1 with x_{il} being the original observation. Therefore, $z_{il} = (x_{il} - \bar{x}_l)/sd(x_l)$. b_0 is the standardised intercept term and coefficients $b_1, ..., b_7$ are the standardised coefficients for $x_1, ..., x_7$, whereas b_8 and b_9 correspond to covariate gender and type of diabetes respectively, and b_{10} represents different levels of retinopathy factors.

To simplify notation, we use $w_i = \sigma_i^{-2}$. The generalised linear model response variable is the estimated posterior mean of the precision parameter, $\widetilde{w}_i = \widetilde{E(w_i \mid y_i)}$. We assume

$$\widetilde{w}_i \sim gamma(\lambda, \frac{\lambda}{m_i}) \qquad \text{for i}=1, 2, ..., I, \tag{5.1.2}$$

so we have $E(\widetilde{w}_i) = m_i$ and $\text{var}(\widetilde{w}_i)=\frac{m_i^2}{\lambda}$.

Parameter m_i is the mean consistency response and linked to all patient-specific covariates through function

$$m_i = \exp(\boldsymbol{Z}^*\boldsymbol{b}) \text{ where } i = 1, 2, ..., I. \tag{5.1.3}$$

$\boldsymbol{b} = (b_0, b_1, ..., b_{10,1}, b_{10,2}, b_{10,3})^T$ is a vector of coefficients corresponding to the vector of covariates $\boldsymbol{Z}^* = (1, z_1, z_2, z_3, z_4, z_5, z_6, z_7, GEN, TYPE, RET1, RET2, RET3)$, as shown in detail in Table 5.1.

All b_l, $l = 1, 2, ..., 12$ are assumed to have Normal prior distributions and the prior for λ is inverse-gamma:

$$b_l \sim Normal(\mu_{\beta_l}, \sigma^2_{\beta_l}) \qquad \text{where } \mu_{\beta_l} = 0, \sigma^2_{\beta_l} = 10^4 \tag{5.1.4}$$

$$\lambda \sim Inverse - gamma(\gamma_\lambda, \delta_\lambda) \qquad \text{where } \gamma_\lambda = \delta_\lambda = 10^{-3}. \tag{5.1.5}$$

The likelihood of this model is given by

$$\begin{aligned} f(\tilde{w}_i|\boldsymbol{b}, \lambda) =& \prod_{i=1}^{I} \frac{(\lambda/\exp(\boldsymbol{Z}^*\boldsymbol{b}))^\lambda}{\Gamma(\lambda)} \tilde{w}_i^{\lambda-1} \exp\left\{-\frac{\lambda}{\exp(\boldsymbol{Z}^*\boldsymbol{b})}\tilde{w}_i\right\} \\ =& \frac{\lambda^{I\lambda}}{\{\Gamma(\lambda)\}^I} \prod \left\{[\exp(\boldsymbol{Z}^*\boldsymbol{b})]^{-\lambda} \tilde{w}_i^{\lambda-1}\right\} \exp\left\{-\lambda \sum_{i=1}^{I} \frac{\tilde{w}_i}{\exp(\boldsymbol{Z}^*\boldsymbol{b})}\right\} \end{aligned} \tag{5.1.6}$$

and the joint posterior density is

$$\pi(\lambda, \boldsymbol{b}|y_{ijk}) \propto f(\tilde{w}_i|\lambda, \boldsymbol{b}) p(\lambda) p(\boldsymbol{b}), \tag{5.1.7}$$

where $p(\boldsymbol{b})$ and $p(\lambda)$ are given in 5.1.4 and 5.1.5 respectively.

We note that there are other options when choosing the distribution for the type of model fitted here. For example, a model with log-normal errors is commonly used. However, fitting a log-normal model does not provide a better fit when compared to the gamma error model (Zammitt et al., 2011).

There are missing values in the data where eight subjects have unspecified records for covariates retinopathy, C-peptide and A1c haemoglobin. This is a common problem in research and can be overcome using the Bayesian paradigm. MCMC techniques are used to impute values for the missing covariates where the covariates are treated as random variables, allowing their posterior distributions of the incomplete data given the observed data to be estimated by the model (Ibrahim et al., 2002). Covariates C-peptide and A1c haemoglobin are continuous measurements and can be denoted by $x_{4,mis}$ and $x_{5,mis}$ respectively where x_{mis} refers to the missing value. When treated as random variable, a probability distribution must be specified to these covariates which will give rise to the values of covariates. We assumed these covariates to be

normally distributed

$$z_4 \sim Normal(\mu_4, \tau_4^2),$$

$$z_5 \sim Normal(\mu_5, \tau_5^2),$$

and the prior distributions assigned are

$$\mu_4, \mu_5 \sim Normal(0, 10^3),$$

$$\tau_4, \tau_5 \sim Inverse - gamma(1, 1).$$

The effect of each covariate is assessed using 95% equal-tailed Bayesian Intervals of their corresponding b coefficients.

From (5.1.7), we get the estimates of the mean and standard deviation of each b coefficient corresponding to covariates. Table 5.2 lists these posterior estimates which were obtained from 20,000 iterations after 5000 burn-in period. As can be noticed in Figure 5.2, all chains converge very well for all b_l parameters in the sampled iterations. A caterpillar plot representing the posterior estimates with their 95% credible intervals is given in Figure 5.3. Covariates gender and no retinopathy have significant effects on consistency. This agrees with the findings in Zammit et al. (2011) where it was found that gender and awareness are two factors affecting consistency. However, after removing patients with highest score of awareness from their analysis, only gender was significant. The mean of the gender coefficient, b_8, is -0.5417 implying that female subjects are less consistent than male subjects in reporting their symptoms of hypoglycaemia. Although covariate awareness of hypoglycaemia now appears to be not significant, this covariate is marginal with $b_7 = 0.0988$ and 95% credible interval $(-0.0293, 0.2278)$. The posterior mean of coefficient $b_{10,1}$, corresponding to 'no retinopathy' is 0.4728 with 95% credible interval (0.0194, 0.9077) meaning that patients with no retinopathy are more consistent in reporting their hypoglycaemic symptoms in contrast to those who have retinopathy.

Parameter	Mean	SD	2.5% CI	97.5% CI
b_1	-0.0062	0.1219	-0.2449	0.2320
b_2	0.0994	0.1720	-0.2470	0.4349
b_3	0.1207	0.1066	-0.0810	0.3423
b_4	-0.1567	0.1199	-0.3963	0.0836
b_5	-0.1512	0.1089	-0.3383	0.0919
b_6	-0.1483	0.1126	-0.3608	0.0775
b_7	0.0988	0.0640	-0.0293	0.2278
b_8	-0.5417	0.2265	-0.9872	-0.0886
b_9	0.5190	0.3203	-0.1099	1.1460
$b_{10,1}$	0.4728	0.2226	0.0194	0.9077
$b_{10,2}$	-0.2743	0.1553	-0.5808	0.0341
$b_{10,3}$	-0.1985	0.2502	-0.6741	0.3073

Table 5.2: Posterior estimates of b coefficients corresponding to patient-specific covariates for model with grouped symptoms.

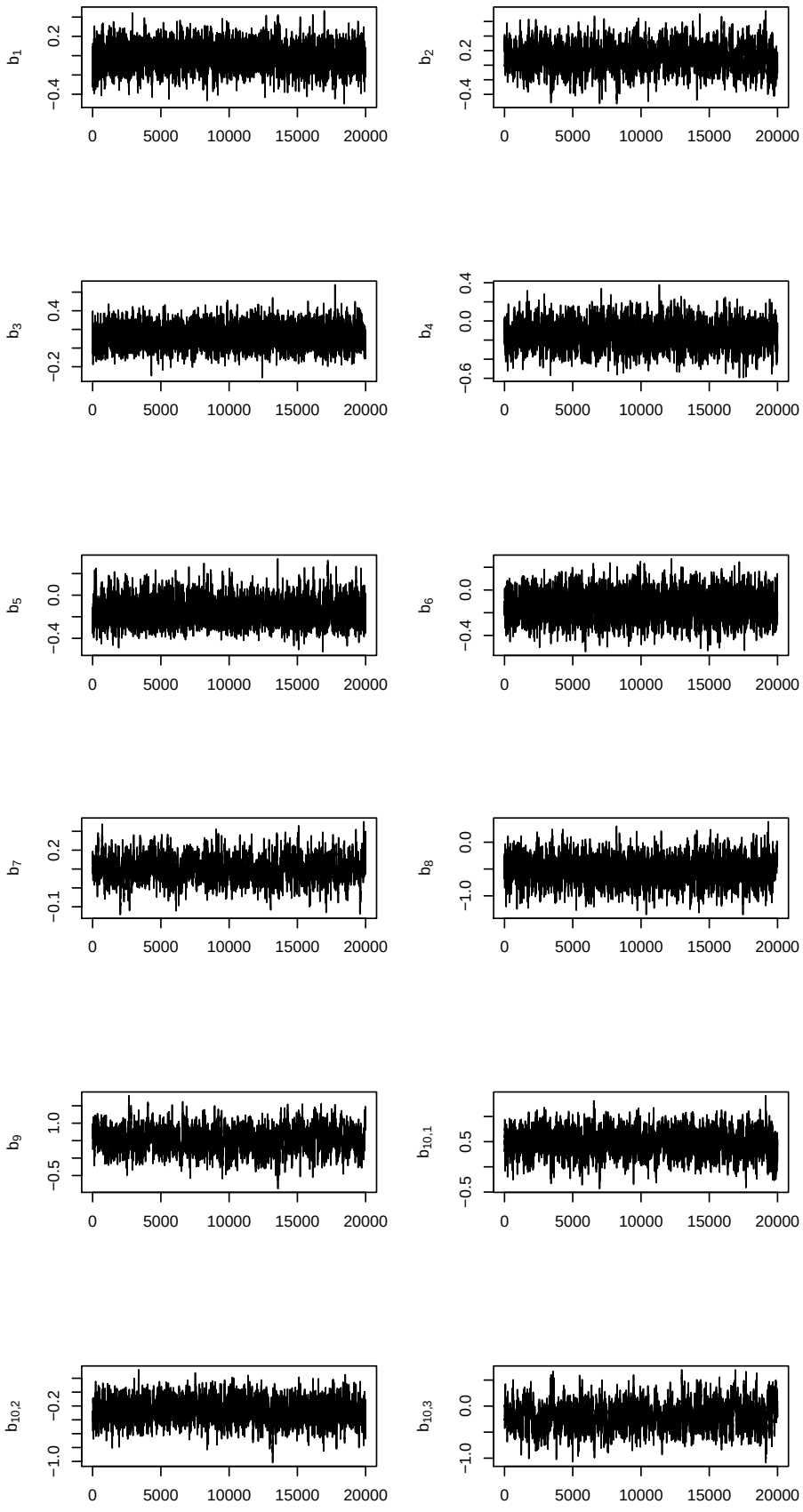


Figure 5.2: Trace plots of b_l coefficients in model with grouped symptoms.

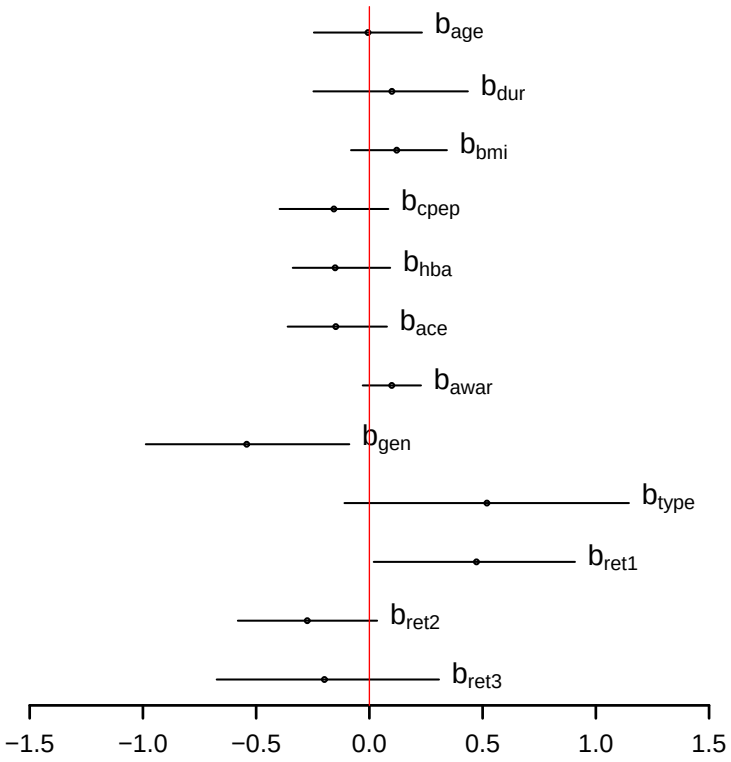


Figure 5.3: Posterior means (bullets) and 95% equal-tailed Bayesian intervals (bars) for standardised coefficients of patient-specific covariates for model with grouped symptoms.

5.1.1 Effect of interactions between covariates

We are also interested to investigate how the interactions between the patient covariates affect the consistency. We would like to see whether the effect on consistency of one factor is not the same at all levels of another factor. For example, unequal levels of consistency between Type I and Type 2 patients might be much different for males than for females. We will consider a number of interaction terms. Gender is a significant factor affecting consistency, and therefore we want to explore possible interactions with other covariates.

Based on unpublished expert information, there are indications that duration of diabetes is a factor that can potentially interact most with the other factors. Patients with longer duration of diabetes will tend to have a higher prevalence of retinopathy. Duration is also associated with diminishing awareness of hypoglycaemia symptoms. Other than that, duration may also be associated with lower levels of C-peptide. C-peptide is a C shaped protein that is cleaved off the proinsulin molecule to make active insulin. It is a marker of endogenous insulin production and exists with insulin in a 1:1 molar ratio, that is one C-peptide molecule produced for every insulin molecule produced. There is no C-peptide in insulin injections. Therefore, it is a good way of telling how much insulin a person is making for themselves, even when they are on insulin injections as a top up.

C-peptide can also possibly interact with type of diabetes. Patients with Type 1 diabetes lose insulin production completely within a few years of diagnosis whereas those with Type 2 diabetes lose it much more gradually over the years.

Awareness of diabetes has the potential to interact with age as it is known that older patients experience less intense symptoms of hypoglycaemia than younger patients.

Another possible interaction is between awareness and type of diabetes. Impaired awareness of diabetes tends to be more common in Type 1 diabetes of long duration than in diabetes of short duration. It can also develop in long-standing Type 2 diabetes but less frequently than in Type 1. Thus, awareness may also interact with

duration of diabetes.

We therefore have a new linear predictor, replacing (5.1.1), given as

$$\begin{aligned}\log\{\tilde{w}_i\} = \log\{m_i\} = b_0 + \sum_{l=1}^{7} b_l z_{il} + b_8 GEN_i + b_9 TYPE_i + b_{10,1} RET_1 \\ + b_{10,2} RET_2 + b_{10,3} RET_3 + b_{gen\times type} \times (GEN.TYPE)_i \\ + b_{gen\times dur} \times (GEN.DUR)_i + b_{gen\times awar} \times (GEN.AWAR)_i \\ + b_{gen\times bmi} \times (GEN.BMI)_i + b_{cpep\times dur} \times (CPEP.DUR)_i \\ + b_{cpep\times type} \times (CPEP.TYPE)_i + b_{awar\times age} \times (AWAR.AGE)_i \\ + b_{awar\times dur} \times (AWAR.DUR)_i + b_{awar\times type} \times (AWAR.TYPE)_i \\ + b_{dur\times ret1} \times (DUR.RET1)_i + b_{dur\times ret2} \times (DUR.RET2)_i \\ + b_{dur\times ret3} \times (DUR.RET3)_i + b_{gen\times ret1} \times (GEN.RET1)_i \\ + b_{gen\times ret2} \times (GEN.RET2)_i + b_{gen\times ret3} \times (GEN.RET3)_i \\ i = 1, 2, ..., 66 \end{aligned} \tag{5.1.8}$$

As previously, we assign a normal prior distribution to b_l coefficients where $l = 1, ..., 27$ such that:

$$b_l \sim Normal(0, 10^3).$$

We sample the posterior coefficient estimates after 15,000 MCMC iterations where the first 5000 are the discarded burn-in iterations. The trace plots we monitored show that convergence had been achieved after the burn-in period (refer Figures 5.4-5.5). A full list of summary statistics of the covariates and their interactions is provided in Table 5.3. Figure 5.6(a) plots the 95% credible intervals of the posterior estimates corresponding to the main covariates and Figure 5.6(b) gives a similar plot representing the interactions between covariates. The mean for gender coefficient was positive, $b_8 =$

1.6250 (95% CI 0.6432, 2.7430). If we take into account the overall effect of gender, the estimated consistency of female patients (mean = 48.51, 95% CI (23.27,78.70)) is lower than that of male patients (mean = 72.39, 95% CI (50.84, 88.83)). The ratio of male to female consistency is 1.63 (95% CI (0.9066, 2.9500)). It is interesting to see that gender, which is highly significant in the previous analysis now shows not to be significant. We suspect this is the effect of including gender in other interaction terms thus pulling or stretching the effect that gender has.

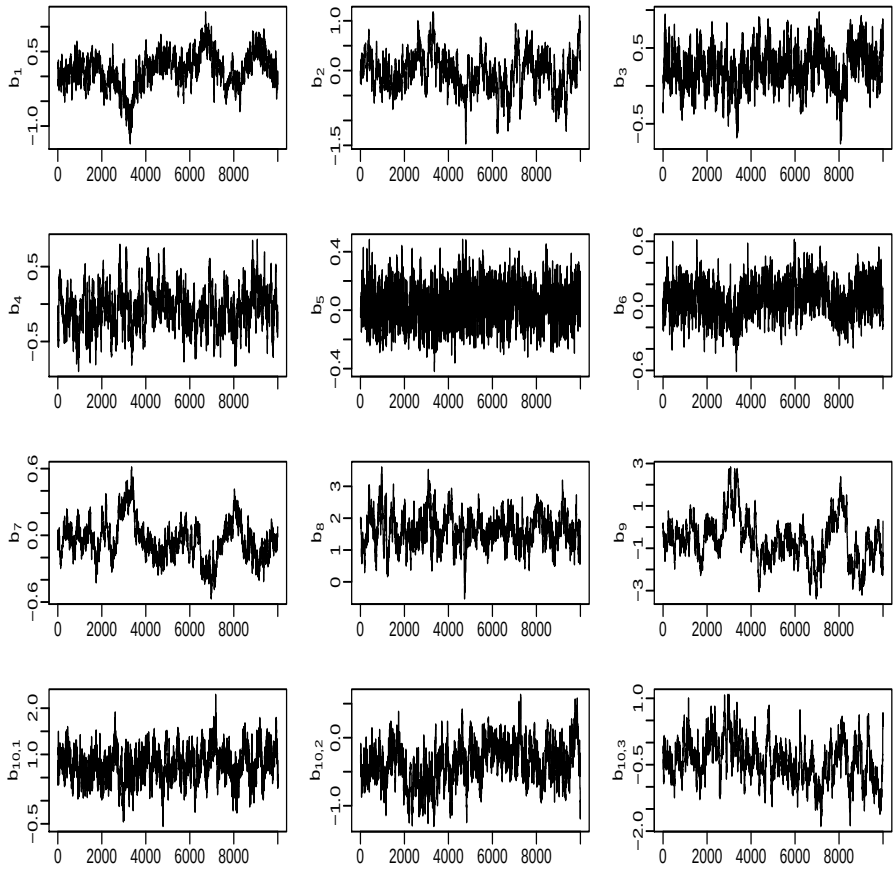


Figure 5.4: Trace plots of b coefficients corresponding to the main covariates when interaction between covariates are included.

The significant estimates for coefficient $b_{10,1}$ show that lack of retinopathy also affects the consistency measure. The estimate of $b_{10,1}$ is 0.7806 (95% CI (0.1069, 1.4590)) which tells us that patients with no retinopathy recorded lower variability in their symptoms than those with mean level of retinopathy.

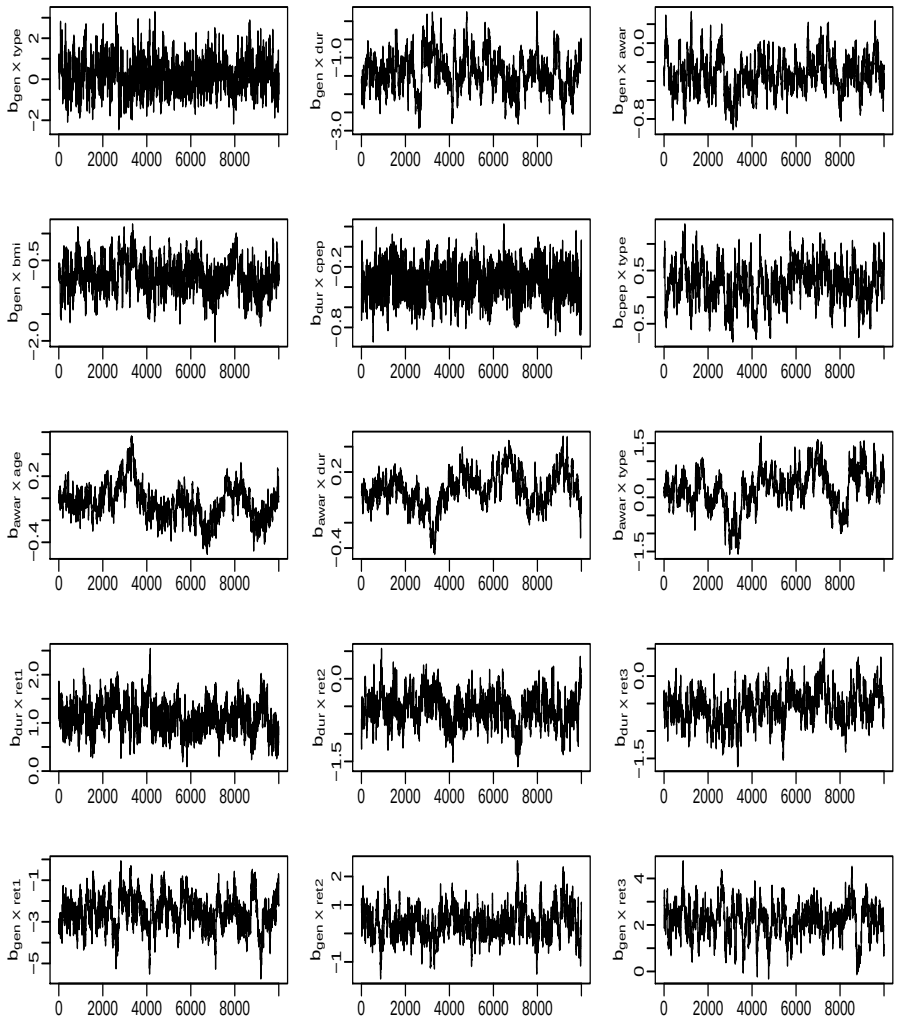


Figure 5.5: Trace plots of b coefficients corresponding to the interaction terms between covariates.

This analysis suggests a number of significant interactions involving covariate gender. These include interactions between gender with duration of diabetes, BMI, no retinopathy, and proliferative retinopathy. $b_{genXdur}$ is -1.404 with 95% credible interval $(-2.429, -0.366)$ indicating that for female patients, as the duration of diabetes increases, their consistency of reporting symptoms decreases. Consistency of female patients also reduced with higher BMI, $b_{genXbmi} = -0.8185$ and 95% credible interval $(-1.417, -0.1984)$.

Also, the results indicate that the interaction between covariate gender and 'no retinopathy' is significant with $b_{genXret1} = -2.5590$ and 95% credible interval $(-4.161,$

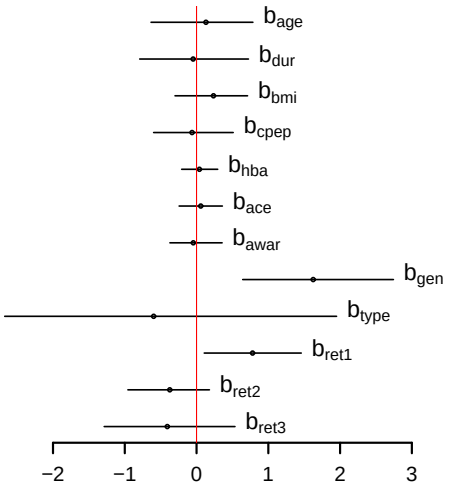
Parameter	Mean	SD	2.5% CI	97.5% CI
b_1	0.1309	0.3438	−0.6343	0.7852
b_2	-0.04742	0.3772	−0.7921	0.7234
b_3	0.2360	0.2554	−0.3002	0.7108
b_4	-0.0627	0.2767	−0.5981	0.5105
b_5	0.0400	0.1301	−0.2075	0.2949
b_6	0.0585	0.1563	−0.2429	0.3600
b_7	−0.0452	0.1753	−0.3712	0.3563
b_8	1.6250	0.5260	0.6432	2.7430
b_9	−0.5976	1.0800	−2.6730	1.9510
$b_{10,1}$	0.7806	0.3489	0.1069	1.4590
$b_{10,2}$	−0.3728	0.2999	−0.9566	0.1794
$b_{10,3}$	−0.4077	0.4622	−1.2870	0.5355
$b_{genXtype}$	0.1694	0.8235	−1.3150	1.9320
$b_{genXdur}$	−1.4040	0.5135	−2.4290	−0.3660
$b_{genXawar}$	−0.3339	0.1881	−0.7049	0.0370
$b_{genXbmi}$	−0.8185	0.3052	−1.4170	−0.1984
$b_{durXcpep}$	−0.3536	0.1556	−0.6600	−0.0550
$b_{cpepXtype}$	0.2467	0.3597	−0.4993	0.9115
$b_{awarXage}$	−0.0379	0.1593	−0.3379	0.3277
$b_{awarXdur}$	0.0633	0.1418	−0.2290	0.3360
$b_{awarXtype}$	0.2195	0.5508	−1.0290	1.2250
$b_{durXret1}$	1.1120	0.3088	0.5377	1.7450
$b_{durXret2}$	−0.5548	0.2882	−1.1420	0.0260
$b_{durXret3}$	−0.5568	0.3100	−1.1820	0.0855
$b_{genXret1}$	−2.5590	0.7748	−4.1610	−1.095
$b_{genXret2}$	0.3562	0.5373	−0.6855	1.4650
$b_{genXret3}$	2.2030	0.7168	0.6256	3.5280

Table 5.3: Posterior estimates of b coefficients for all covariates and their interactions.

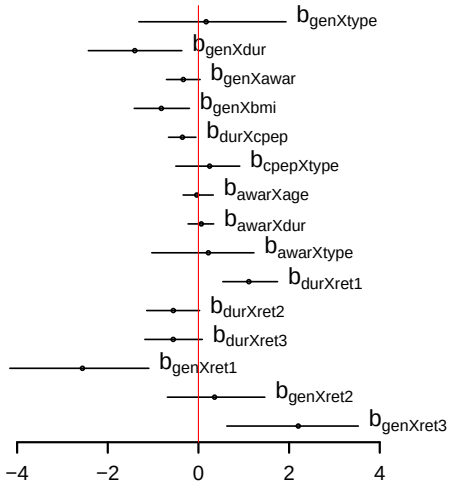
−1.095). This means that female subjects with no retinopathy have lower consistency compared to their male counterpart. However, female subjects with proliferative retinopathy have higher consistency than male. This is given by the posterior estimate, $b_{genXret3} = 2.2030$ with 95% credible interval (0.6256, 3.5280). Note that with different type of retinopathy, female patients have different effect on consistency than male patients.

Another interaction found to be significant in affecting consistency is between covariates duration and no retinopathy. Consistency of symptom reportings rises as duration increases for patients with no retinopathy. The posterior estimate of the coefficient corresponding to this interaction is $b_{durXret1} = 1.1120$ with 95% credible interval

(0.5377, 1.7450). Patients with longer duration of diagnosed diabetes have lower consistency with the increment of C-peptide level. This is given by $b_{durXcpep} = -0.3536$ and 95% credible interval of $(-0.6600, -0.0550)$.



(a) Main factors



(b) Interactions between factors

Figure 5.6: Posterior means (bullets) and 95% equal-tailed Bayesian intervals (bars) for standardised coefficients of patient-specific covariates and their interactions.

5.2 Hierarchical model

In Chapters 2 and 3 we explain the model for consistency estimation, whereas early sections of this chapter discuss the GLM methodology for checking the association between covariates and consistency. In this section, we propose a hierarchical model, where, unlike in previous work where it was needed to obtain the precision parameter first before using the posterior estimation in GLM to test the covariates effect on consistency, this hierarchical model does both estimations in one setting.

Referring to the consistency model in Section 2.5, we have

$$\begin{aligned} y_{ijk} \sim & Bernoulli(p_{ijk}), \\ p_{ijk} = & Pr\{\tau_{ijk} \leq \alpha_{ij}\beta_{ik}\}, \\ = & \Phi\left\{\frac{\log[\alpha_{ij}\beta_{ik}]}{\sigma_i}\right\}, \end{aligned}$$

where i = 1,...,66 patients, j = 1,...,26 symptoms and k = 1,...,K_i episodes. y_{ijk} is the observation variable and p_{ijk} is the probability of a threshold being less than or equal to latent variables α_{ij} and β_{ik}, with α_{ij} being the propensity of symptoms j for patient i and β_{ik} the intensity of episode k for patient i.

As before,

$$\tau_{ijk} \sim LN(0, \sigma_i^2).$$

To allow for hierarchical estimation, the variance parameter, σ_i^2, follows a hierarchical inverse gamma distribution:

$$\sigma_i^2 \sim \text{Inv-gamma}(\delta, \delta/m_i)$$

where

$$\delta \sim \text{Inv-gamma}(a_\delta, b_\delta).$$

To facilitate the MCMC algorithm's convergence, we set $a_\delta = b_\delta = 1$ as it reduces

the variability of hyper-parameter δ. As defined earlier, parameter m_i is the mean consistency response and linked to each covariate through function $m_i = \exp(\boldsymbol{Z}^*\boldsymbol{b})$ where i = 1,2,...,I and $\boldsymbol{b} = (b_0, b_1, ..., b_p)^T$ is a vector of coefficient corresponding to a covariate vector $\boldsymbol{Z}^* = (1, z_1, z_2, z_3, z_4, z_5, z_6, z_7, GEN, TYPE, RET1, RET2, RET3)$. For this hierarchical model, interactions between covariates are not included. Therefore, $p = 12$. The b coefficients have the same prior as given in (5.1.4).

Since we are using the model with grouped symptoms, the prior for latent variable α_{ijl} is as defined in (3.4.1), that is

$$\alpha_{ijl} \sim Gamma(\theta, \theta/u_l)$$

$$\text{with } \theta \sim Gamma(a_\theta, b_\theta)$$

$$u_l \sim Gamma(a_{u_l}, b_{u_l}).$$

The prior distribution given to latent variable β_{ik} is

$$\beta_{ik} \sim Gamma(a_\beta, b_\beta).$$

We set θ equal to 1 and use $a_{u_l} = b_{u_l} = a_\beta = 1$ and $b_\beta = 0.1$.

We observe the trace plot of the model's parameters to check their convergence (Figures 5.7- 5.9). Their posterior distributions were sampled after a 5000 burn-in period. It is obvious from the plots that chains of parameter σ_i^2 have achieved stationarity. However, we only present trace plots for twelve subjects; 5009, 3057, 3065, 1055, 3048, 3052, 3067, 4076, 3016, 6002, 4072, and 6056 in Figure 5.7. Figure 5.8 shows the sample paths of the simulation for parameter $\boldsymbol{b}$. Again, the chains are strongly converging without any problems. This is also true for the group's propensity parameter, u_l (Figure 5.9).

Consistency estimates for all patients obtained from this model are comparable with what we obtained with the non-hierarchical consistency estimate model with grouped symptoms in Section 3.4. Figure 5.10 (left) gives the distribution of the posterior

mean estimates of the consistency parameter, $\tilde{c}_i$, for all 66 patients in the hierarchical model with grouped symptoms, whereas Figure 5.10 (right) shows the estimates of $\log(\sigma_i^{-2})$. Subject 1028 was the most consistent patient among all 66 patients when we estimated the consistency using the non-hierarchical model. However, when using the hierarchical model, this subject has become the second most consistent patient, with $\tilde{c}_i = 80.87$. The least consistent patient among all is still subject 2013, $\tilde{c}_i = 11.53$ (marked by red lines in Figure 5.10 (left)). Figure 5.11 gives the rank of consistency estimates for this hierarchical model versus the non-hierarchical model. Patients are ranked such that the patient with the highest consistency is ranked first and patient with the lowest consistency is ranked last. This plot shows a decent correspondence suggesting that the estimations obtained from the hierarchical and non-hierarchical model are similar. Posterior estimates of $\tilde{c}_i$ for all patients are given in Table 5.4.

Table 5.5 provides posterior estimates of the mean propensity in each group, u_l for $l = 1,...,6$, and their distributions are shown in Figure 5.12. The distributions are very similar to what we get when the non-hierarchical model is used, as can be observed from the violin plot in Figure 5.13. Autonomic and neuroglycopenic groups have the highest propensity suggesting their importance in detecting the onset of hypoglycaemia ($u_1 = 0.0420$ and $u_2 = 0.0377$ respectively).

We can also investigate the covariates' effect using this hierarchical model. Figure 5.14 shows the caterpillar plots for coefficients $\boldsymbol{b}$ corresponding to each covariate, and the summary statistics of posterior estimates of $\boldsymbol{b}$ are given in Table 5.6. Recall that for the non-hierarchical model with grouped symptoms in Section 5, covariates gender and no retinopathy significantly affected the consistency reporting (Figure 5.3). However, the implementation of hierarchical modeling to grouped symptoms results in no covariate appearing to be affecting significantly the consistency, as exhibited by the 95% credible intervals of all b coefficients containing zero (See Figure 5.14). The hierarchical setting allows for more variation about the consistency-related parameter (σ_i^2) to enter the model (through the Inverse-gamma distribution), and this can potentially account for differences in patient characteristics (covariates) not being identified. Figure 5.15 displays the contrast between posterior densities of the coef-

ficients for covariates gender and no retinopathy in non-hierarchical and hierarchical model. Under the hierarchical model, the density of gender seems to have shifted to the right whereas the density of no retinopathy has shifted to the left.

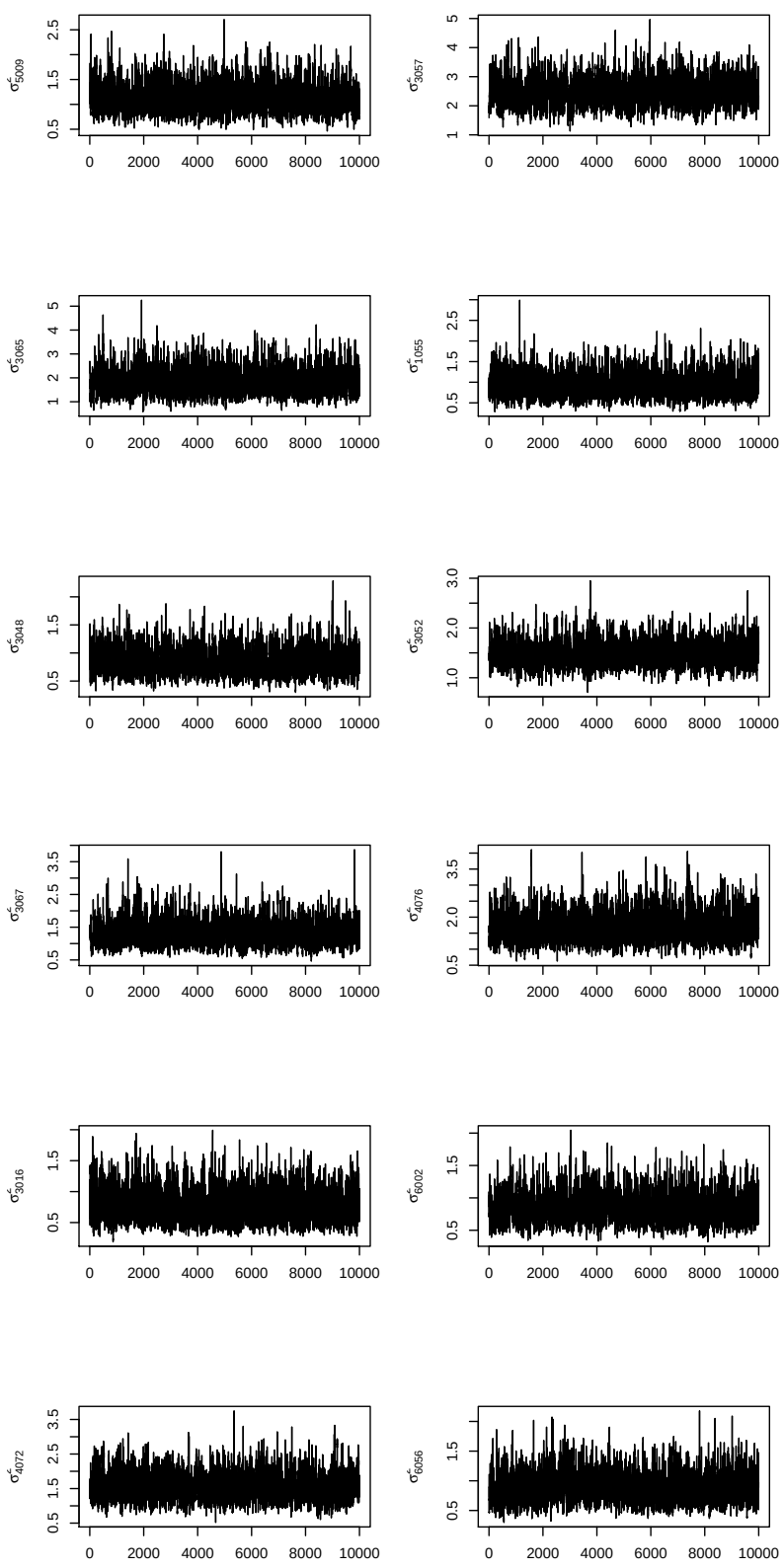


Figure 5.7: Trace plots of σ_i^2 coefficients in hierarchical model with grouped symptoms, for selected patients i.

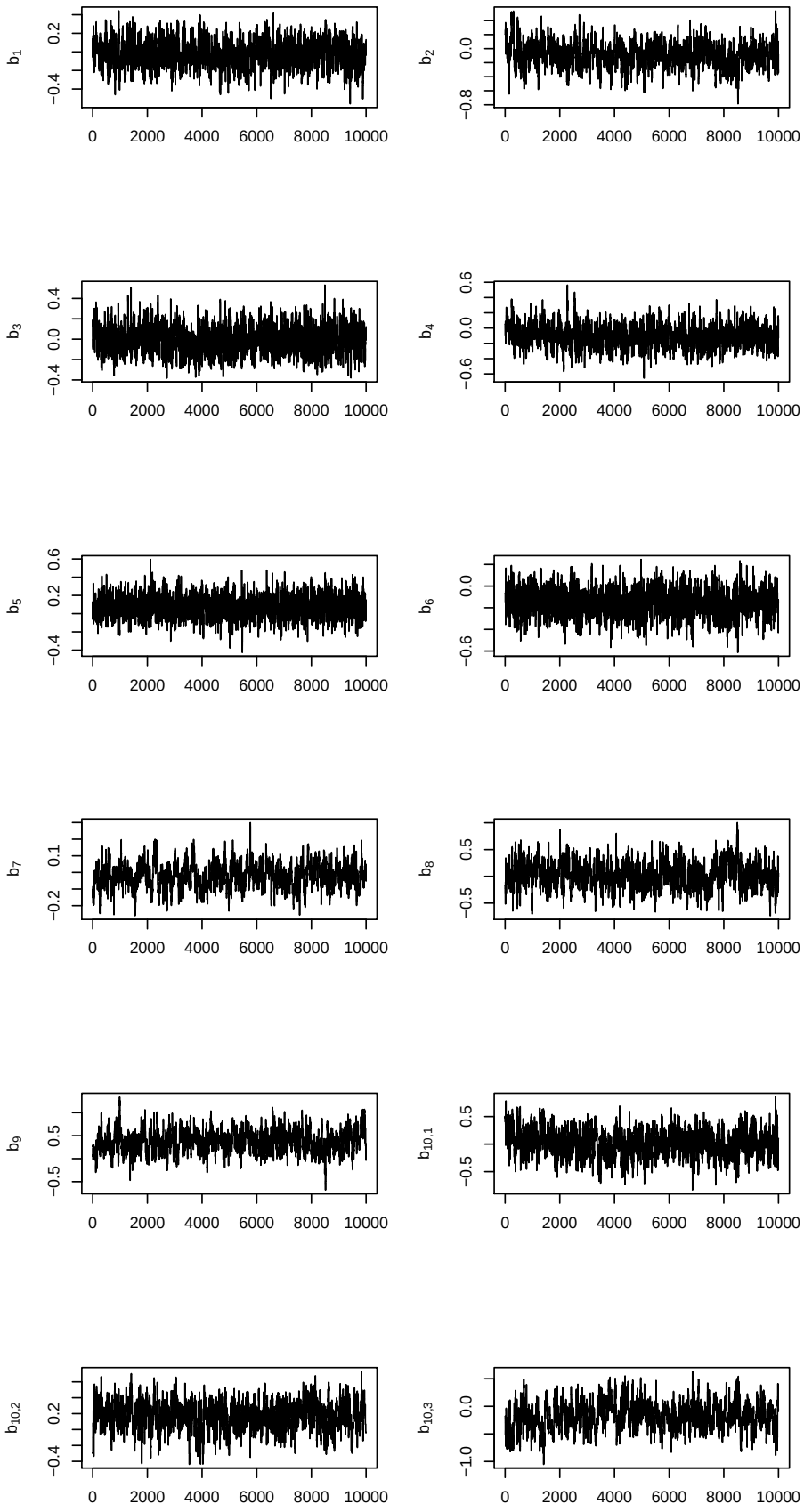


Figure 5.8: Trace plots of $\boldsymbol{b}$ in hierarchical model with grouped symptoms.

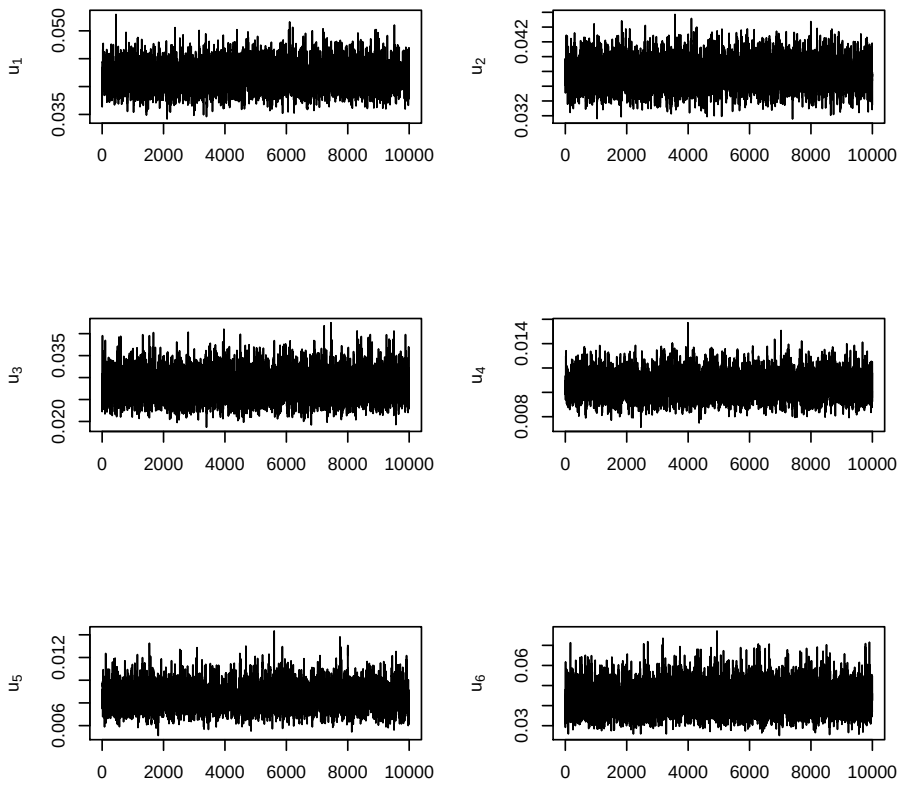


Figure 5.9: Trace plots of group parameter, u_l for $l = 1,...6$ where u_1 refers to Autonomic, u_2 to Neuroglycopenic, u_3 to Autonomic/Neuroglycopenic, u_4 to General malaise, u_5 to Other and u_6 to No symptom in hierarchical model with grouped symptoms.

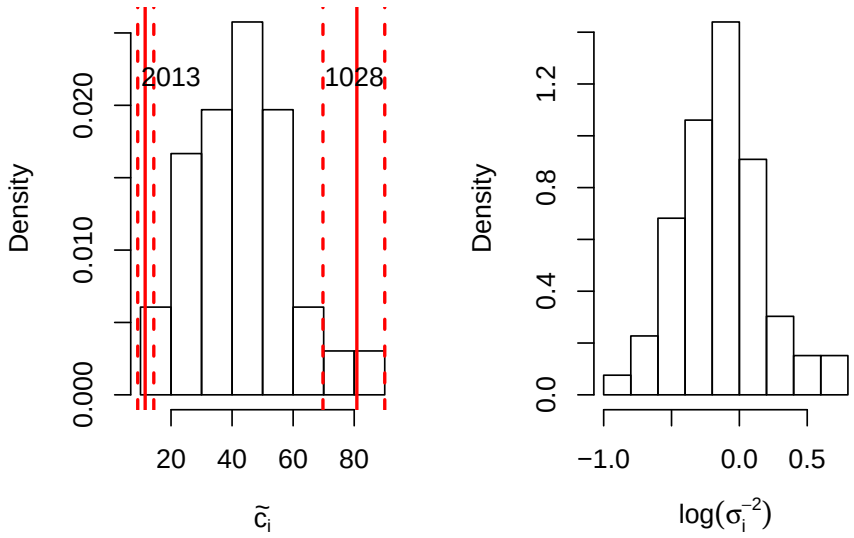


Figure 5.10: Histogram of estimated consistency parameter, $\tilde{c}_i$ and the log estimated precision parameter, σ_i^{-2} for hierarchical model with grouped symptoms.

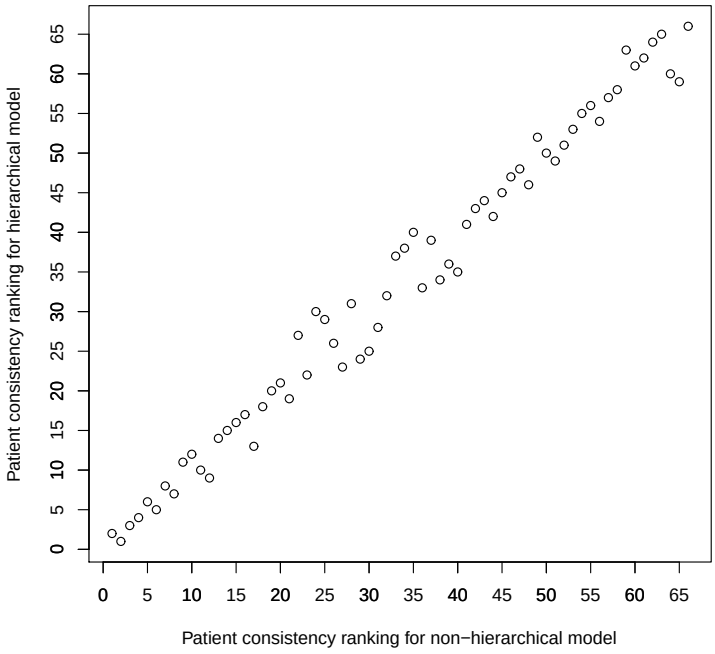


Figure 5.11: Graph shows ascending order for ranking of consistency estimates, $\tilde{c}_i =$ of 66 patients for hierarchical and non-hierarchical model for model with grouped symptoms.

Subject	No. of episode	$\tilde{c}_i$	SD	25% CI	95% CI
1008	123	26.20	3.220	20.33	32.83
1009	96	47.50	4.573	38.86	56.66
1015	34	32.82	5.803	22.39	44.98
1021	49	41.91	5.731	31.30	53.84
1025	26	39.73	6.809	27.36	53.63
1028	45	80.87	5.232	69.77	90.01
1036	42	49.83	6.162	38.07	62.01
1039	47	31.88	5.282	22.37	42.96
1055	27	53.31	7.104	39.53	67.25
1057	11	75.91	7.012	60.66	87.73
1086	91	27.36	3.771	20.38	35.24
2009	101	24.00	3.242	18.16	30.74
2010	102	25.29	3.210	19.39	32.06
2012	42	27.21	4.736	18.61	37.18
2013	300	11.53	1.331	9.09	14.36
2015	44	33.09	4.847	24.43	43.20
2021	32	39.96	6.133	28.57	52.11
2022	41	36.54	5.360	26.67	47.82
2027	26	44.92	6.950	31.61	58.89
3001	35	46.66	6.495	34.24	59.45
3015	19	15.35	2.908	10.46	21.80
3016	23	58.53	7.314	43.86	72.55
3022	23	46.42	7.489	32.39	61.85
3024	30	54.54	6.658	41.71	67.76
3029	78	69.87	4.913	59.55	78.91
3043	17	41.80	8.273	26.19	58.55
3046	25	45.60	6.944	32.64	59.84
3048	37	55.07	6.263	42.87	67.24
3050	54	51.89	5.171	41.67	61.93
3052	146	40.82	3.902	33.24	48.80
3057	91	29.45	3.839	22.38	37.51
3065	27	36.73	6.388	25.19	50.12
3067	32	44.22	6.331	32.08	56.97
4003	26	55.81	6.800	42.40	68.81
4008	69	48.51	4.635	39.56	57.85
4013	55	30.18	4.532	22.19	39.78
4023	211	20.69	2.241	16.50	25.41
4028	12	51.42	9.291	32.63	68.69
4032	10	65.42	9.546	45.53	82.37
4034	61	63.24	4.986	53.12	72.41
4043	36	30.43	5.217	21.01	41.25
4045	47	32.07	4.863	22.98	42.09
4049	23	54.10	7.046	40.05	67.79
4061	12	74.42	7.184	58.95	86.89
4063	48	27.37	4.429	19.37	36.59
4072	40	40.27	5.729	29.51	51.95
4076	36	38.45	5.935	27.52	50.44
5004	28	40.23	6.332	28.74	53.26
5009	40	48.14	5.808	36.89	59.63
5023	22	47.33	7.725	33.01	62.76
5026	26	57.55	6.886	44.03	70.93
5029	47	58.82	5.905	47.14	70.05
5044	87	29.56	3.744	22.67	37.35
5045	78	40.68	4.640	32.12	50.13
5088	100	18.73	2.657	14.03	24.39
6002	34	55.22	6.098	43.32	67.07
6010	19	51.84	7.786	36.47	66.81
6018	62	37.60	4.701	28.84	47.10
6019	93	30.14	3.777	23.03	37.96
6023	76	24.46	3.562	18.03	31.95
6038	122	24.28	3.056	18.91	30.73
6056	36	54.18	6.494	41.52	66.95
6058	20	62.27	6.973	48.21	75.09
6062	13	48.52	9.195	30.70	66.30
6064	17	82.79	5.010	71.62	91.20
6065	44	19.61	4.491	12.34	29.63

Table 5.4: Posterior estimates of consistency (Mean, standard deviation, and 95% credible interval) for all subjects using hierarchical model with grouped symptoms.

Group	Mean	SD	2.5% CI	97.5% CI
Autonomic	0.0420	0.0023	0.0376	0.0467
Neuroglycopenic	0.0377	0.0019	0.0341	0.0416
Autonomic/Neuroglycopenic	0.0280	0.0031	0.0225	0.0347
General Malaise	0.0106	0.0010	0.0088	0.0126
Other	0.0085	0.0011	0.0066	0.0108
No Symptom	0.0432	0.0068	0.0317	0.0582

Table 5.5: Posterior estimates of group propensity, u_l when using hierarchical model.

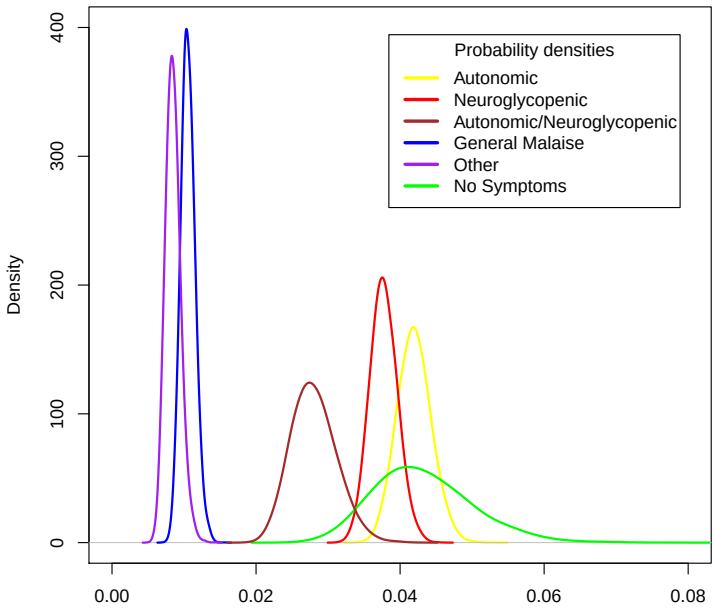


Figure 5.12: Posterior distributions of mean group propensity, u_l in hierarchical model.

Parameter	Mean	SD	2.5% CI	97.5% CI
b_1	−0.0155	0.1355	−0.2787	0.2500
b_2	−0.0936	0.1732	−0.4438	0.2428
b_3	−0.0121	0.1231	−0.2404	0.2440
b_4	−0.1093	0.1390	−0.3778	0.1800
b_5	0.0670	0.1154	−0.1554	0.3001
b_6	−0.1607	0.1175	−0.3913	0.0646
b_7	−0.0235	0.0738	−0.1721	0.1260
b_8	0.0245	0.2363	−0.4232	0.5062
b_9	0.3698	0.2429	−0.0775	0.8732
$b_{10,1}$	0.0135	0.2243	−0.4220	0.4533
$b_{10,2}$	0.1708	0.1673	−0.1809	0.4810
$b_{10,3}$	−0.1843	0.2488	−0.6814	0.3125

Table 5.6: Posterior estimates of b coefficients corresponding to patient-specific covariates for hierarchical model with grouped symptoms.

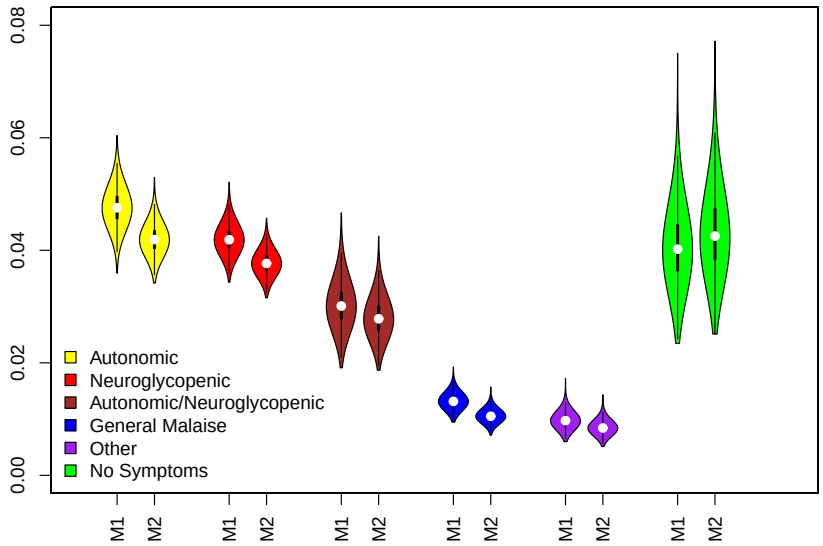


Figure 5.13: Violin plots of posterior distributions of mean group propensity, u_l for $l = 1, ..., 6$ obtained with non-hierarchical model with grouped symptoms (M1) and hierarchical model with grouped symptoms (M2).

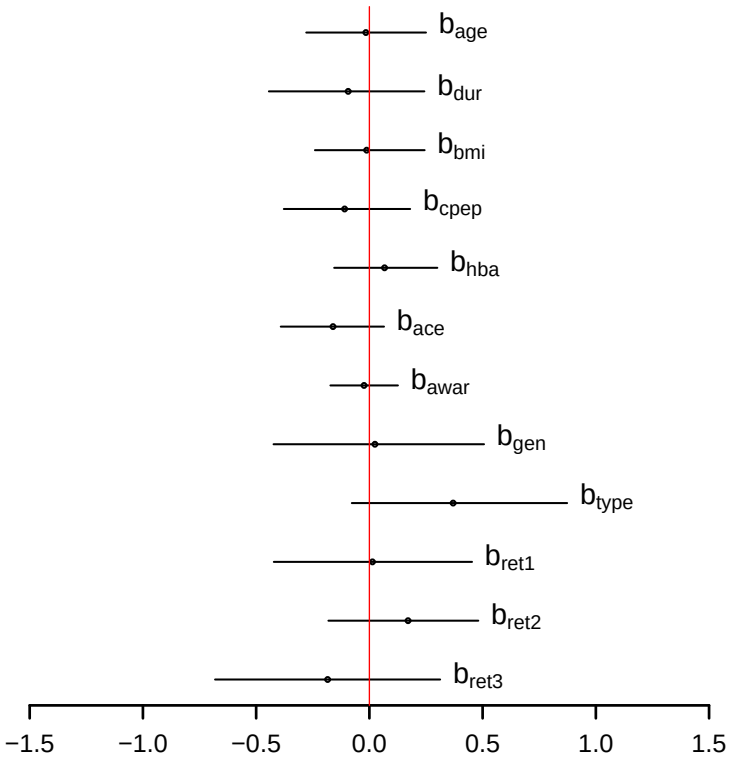


Figure 5.14: Posterior means (bullets) and 95% equal-tailed Bayesian intervals (bars) for standardised coefficients of patient-specific covariates in hierarchical model with grouped symptoms.

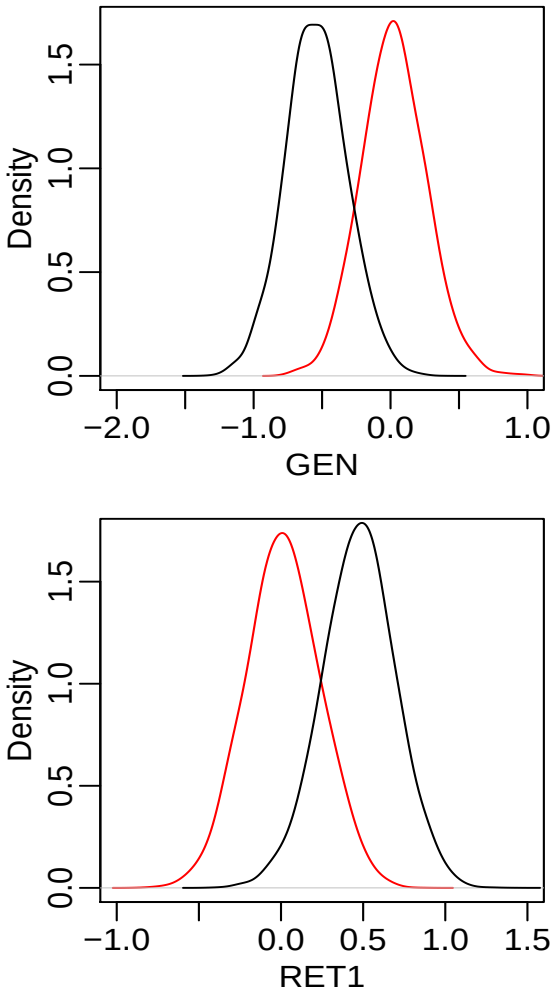


Figure 5.15: Posterior densities of b coefficient for covariate gender (GEN) and no retinopathy (RET1) in hierarchical model and non-hierarchical model with grouped symptoms. The hierarchical model is represented with red lines.

5.3 Concluding remarks

As a conclusion, examining the effect of patient specific covariates on consistency using estimates obtained with the non-hierarchical grouped symptoms model, shows that gender and no retinopathy significantly affect the consistency of symptom reportings for individual patients. When interactions between the covariates are taken into account, three interactions involving gender are significant, i.e. gender $\times$ duration, gender $\times$ BMI, and gender $\times$ retinopathy 1 (no retinopathy). The effect on consistency of covariate duration is also different as the level of C-peptide changes. Another significant interaction is between duration of diabetes and lack of retinopathy.

We have also considered a hierarchical model which is able to estimate patients' consistency and determine which covariates affect it in a single model. The consistency estimates for patients are similar to what we had obtained with the non-hierarchical model in Section 4.3. However, with the hierarchical model, no covariate appears to be significantly affecting patients' consistency.

Chapter 6

Predictive Model Selection

This chapter presents variable selection methodologies used to select the best predictive model for predicting patients' consistency. We first implement a stepwise regression procedure to identify important covariates that should be included in a predictive model. We use this methodology for the model with grouped symptoms, including when we consider interactions between the covariates, and the hierarchical model introduced in Chapter 5. Next, we provide an overview of the second method used, Gibbs variable selection, before discussing the results. In the final section, we use Bayesian model averaging to overcome the model uncertainty issue shown from the Gibbs variable selection.

6.1 Stepwise selection

In Chapter 5, we investigate the association between individuals' consistency and covariates. Using the GLM approach in that chapter, we developed a model for predicting the consistency of symptom reporting among patients. Our aim is to develop a model that efficiently predicts a patient's consistency using the available information of that patient's characteristics. Ideally, we would like to have a simplified model that only consists the best subset of explanatory variables. We take the precision, σ_i^{-2}, in (5.1.1) as the response, since the consistency parameter is $c_i = \dfrac{100}{(1+\sigma_i^2)}$, and the variables listed in Table 5.1 as the predictors. To select the desired model from a large set of plausible models, we use variable selection. One option is to fit all 2^p possible models where p is the number of potential explanatory variables which in

our case is $p = 10$. Therefore we need to consider $2^{10} = 1024$ models and calculate the marginal likelihood estimates for all of them. This is almost impossible and time consuming to be done. Therefore, to avoid actually fitting all the models, we use a stepwise selection procedure.

In general regression-type problems, a stepwise procedure is employed to simplify a model containing a large number of possible covariates to a reduced model containing a subset of 'better' covariates which provide important information about the response variable, thus giving the model a good predictive ability (Miller, 2002). The stepwise selection method adds in or removes one variable at each stage, given that the variable meets a selection criterion for entry or removal, until there is no new predictor that can be added or deleted. The selection criterion used in this analysis is the DIC (as defined in Section 3.3.1). Every time a new variable is added or dropped from the model, the significance of each of the variables already in the model is re-examined and compared sequentially. To run the stepwise regression, we incorporate a binary vector $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{10})$ to the linear predictor in (5.1.1). As defined previously, $w_i = \sigma_i^{-2}$ and $\widetilde{w}_i = \widetilde{E(w_i \mid y_i)}$ is the estimated posterior mean of the precision parameter. So we have:

$$\widetilde{w}_i \sim gamma(\lambda, \frac{\lambda}{m_i}) \qquad \text{for i=1, 2, ..., } I$$

$$\log(m_i) = b_0 + \sum_{q=1}^{p} b_q \gamma_q X_{iq},$$

where $m_i = E(\sigma_i^{-2})$ (Ntzoufras, 2009). We simplify the procedure as follows.

i. Specify the initial model with binary indicator $\boldsymbol{\gamma}^{(0)}$, and set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(0)}$.

ii. The currently selected model is

$$\log(m_i) = b_0 + \sum_{q=1}^{p} b_q \gamma_q X_{iq}.$$

We also have p candidate models characterised by their binary indicators, $\boldsymbol{\gamma}_{\boldsymbol{k}}^{c} =$

$(\gamma^c_{1,k}, ..., \gamma^c_{p,k})^T$ for $k = 1, ..., p$ and each k candidate model is defined as:

$$\log(m^c_{i,k}) = b^c_{0,k} + \sum_{q=1}^{p} b^c_{q,k}\gamma^c_{q,k}X_{iq}$$

where

$$\gamma^c_{q,k} = \gamma_q I(q \neq k) + (1 - \gamma_q)I(q = k),$$

$$b^c_{q,k} = b_q I(q \neq k) + (1 - b_q)I(q = k)$$

with $I()$ being the indicator function. If X_q is included in the model, $\gamma_q = 1$, then it is removed from qth candidate model ($\gamma^c_{q,q} = 1 - \gamma_q = 0$) but is included in all other candidate models ($\gamma^c_{q,k} = \gamma_q = 1$, for all $k \neq q$).

iii. From step ii, we fit p models and estimate their DIC values.

iv. Select the best model among all fitted models, i.e. the model with the lowest DIC value.

v. If the selected model is the same as the one indicated by the previous steps, terminate the procedure. Otherwise, repeat step ii.

Our data set contains missing data as was previously discussed in Chapter 5. For variable selection, we substitute the missing values with their mean posterior estimates (Celeux et al., 2006). Their estimates using the model with grouped symptoms are given in Table 6.1.

Covariate	Mean	SD
$CPEP_{4034}$	0.0056	1.008
HBA_{1028}	−1.2080	1.389
HBA_{1039}	0.1474	1.037
HBA_{2021}	0.0504	1.031
HBA_{3050}	0.1002	1.005
HBA_{4023}	0.2722	1.052
HBA_{6019}	0.0929	1.018

Table 6.1: Posterior estimates of missing values for grouped symptoms model.

We explore three different ways of starting the stepwise procedure which are: 1) starting with the null model, i.e. the model containing only the intercept parameter without any covariate; 2) starting with the full model, i.e. model containing all p covariates. This starting point is suitable to use if $(p + 2 < n)$, where n is the sample size; and 3) starting with model containing only significant covariates (Section 5.1).

Starting the procedure using the null model takes seven cycles before it is terminated, thus 71 models were fit in total. The DIC value from each step is summarised in Table 6.2. First, covariate retinopathy (x_{10}) was added to the null model followed by covariate BMI (x_3). Gender (x_8) and haemoglobin A1c (Hba) (x_5) were added in the third and fourth cycle respectively. Finally, covariate awareness of hypoglycaemia also enters the final predictive model. The model with these covariates is found to be the best model, indicated by its DIC value, 159.13. Therefore, the best predictive model has linear predictor;

$$\begin{aligned}\log\{E(\sigma_i^{-2})\} =& b_0 + b_3 \times BMI_i + b_5 \times HBA_i + b_7 \times AWAR_i + b_8 \times GEN_i \\ &+ b_9 \times TYPE_i + b_{10} \times RET_i \\ i =& 1, 2, ..., 66\end{aligned}$$

with the posterior standardised intercept term, $b_0 = -0.1552$ with 95% CI $(-0.6074, 0.3212)$ and the posterior estimates for coefficients corresponding to each covariates

being

$$b_3 = 0.2194 \text{ with } 95\% \text{ CI } (-0.0049, 0.4680),$$

$$b_5 = -0.1970 \text{ with } 95\% \text{ CI } (-0.3816, -0.0112),$$

$$b_7 = 0.1293 \text{ with } 95\% \text{ CI } (0.0095, 0.2486),$$

$$b_8 = -0.4086 \text{ with } 95\% \text{ CI } (-0.8588, 0.0550),$$

$$b_9 = 0.5010 \text{ with } 95\% \text{ CI } (-0.0397, 1.0670),$$

$$b_{10,1} = 0.4438 \text{ with } 95\% \text{ CI } (0.1306, 0.7531),$$

$$b_{10,2} = -0.2895 \text{ with } 95\% \text{ CI } (-0.6255, 0.0745),$$

$$b_{10,3} = -0.1543 \text{ with } 95\% \text{ CI } (-0.5535, 0.2876).$$

Step	(1)	(2)	(3)	(4)
Previous model: γ	$-$	0000000000	0000000001	0010000001
Action	Initialise	$+x_{10}$	$+x_3$	$+x_8$
Current model DIC	183.87	174.56	168.43	162.65
γ	0000000000	0000000001	0010000001	0010000101
x_1	185.78	174.86	168.92	163.83
x_2	185.44	174.05	169.47	162.80
x_3	177.86	**168.16**	$(*)$174.66	$(*)$169.03
x_4	185.25	176.03	169.49	163.16
x_5	176.90	171.28	165.02	**160.67**
x_6	179.48	171.23	167.60	163.67
x_7	180.00	170.78	166.10	161.73
x_8	179.34	169.05	**162.89**	$(*)$168.44
x_9	183.81	175.42	170.02	164.67
x_{10}	**174.49**	$(*)$183.87	$(*)$177.83	$(*)$173.20

Step	(5)	(6)	(7)
Previous model: γ	0010000101	0010100101	0010101101
Action	$+x_5$	$+x_7$	$+x_9$
Current model DIC	160.79	160.46	159.13
γ	0010100101	0010101101	0010101111
x_1	161.70	160.68	160.62
x_2	161.15	162.59	159.84
x_3	$(*)$167.46	$(*)$166.86	$(*)$160.96
x_4	161.37	162.42	159.22
x_5	$(*)$162.36	$(*)$161.65	$(*)$161.84
x_6	161.06	160.22	159.19
x_7	**160.51**	$(*)$160.79	$(*)$161.88
x_8	$(*)$165.10	$(*)$163.57	$(*)$160.42
x_9	162.05	**159.15**	$(*)$160.35
x_{10}	$(*)$169.47	$(*)$170.48	$(*)$164.96

Key: $(+)$ = add a covariate to the current model, $(-)$ = remove a covariate from a current model, and $(*)$ = covariate present in the current model.

Table 6.2: Tabulated summary of stepwise procedure for grouped symptoms model, starting from null model. Bold values refer to the lowest DIC values in each step.

Using the full model to initiate the stepwise procedure gives a slightly different predictive model which was achieved after three cycles and fitting 31 models. The procedure is summarised in Table 6.3 (a). Covariate age (x_2) was eliminated in the first cycle. Later in the following cycles, covariates duration of diabetes (x_1) was removed before we reach the final model which has the following linear predictor

$$\begin{aligned}\log\{E(\sigma_i^{-2})\} =& b_0 + b_3 \times BMI_i + b_4 \times CPEP_i + b_5 \times HBA_i + b_6 \times ACE_i \\ & b_7 \times AWAR_i + b_8 \times GEN_i + b_9 \times TYPE_i + b_{10} \times RET_i \\ i =& 1, 2, ..., 66\end{aligned}$$

Notice that this model has two extra predictors in comparison to the predictive model obtained when we start with constant model. This model has DIC value of 158.95 and the posterior standardised intercept term, $b_0 = -0.1919$ with 95% CI (−0.6365,0.2899). The posterior estimates of each coefficient are as follows

$$b_3 = 0.2031 \text{ with } 95\% \text{ CI } (-0.0291, 0.4534),$$

$$b_4 = -0.1851 \text{ with } 95\% \text{ CI } (-0.4430, 0.0697),$$

$$b_5 = -0.2065 \text{ with } 95\% \text{ CI } (-0.3876, -0.0193),$$

$$b_6 = -0.1592 \text{ with } 95\% \text{ CI } (-0.3652, 0.0511)$$

$$b_7 = 0.1112 \text{ with } 95\% \text{ CI } (-0.0144, 0.2399),$$

$$b_8 = -0.4606 \text{ with } 95\% \text{ CI } (-0.9405, 0.0277),$$

$$b_9 = 0.7212 \text{ with } 95\% \text{ CI } (0.10501.3710),$$

$$b_{10,1} = 0.4133 \text{ with } 95\% \text{ CI } (0.0995, 0.7200),$$

$$b_{10,1} = -0.2971 \text{ with } 95\% \text{ CI } (-0.6352, 0.0496),$$

$$b_{10,1} = -0.1162 \text{ with } 95\% \text{ CI } (-0.4973, 0.3261).$$

Finally, if we start the procedure with the model containing the only covariate that significantly affects individual consistency, i.e. gender (from Section 5.1), predictor retinopathy (x_{10}) was added in the first cycle. Then, BMI (x_3) and Hba (x_5) were selected as prominent predictors according to their DIC values. During the fourth and fifth cycle, we add awareness of hypoglycaemia (x_7) and type of diabetes (x_9) to the model respectively. The procedure stopped in the sixth cycle giving the final model exactly as what we obtain when using null model as starting points. Table 6.3 (b) gives the summary of this procedure.

Step	(1)	(2)	(3)
Previous model: γ	–	1111111111	1011111111
Action	Initialise	$-x_2$	$-x_1$
Current model DIC	163.05	160.64	158.95
γ	1111111111	1011111111	0011111111
x_1	(∗)161.09	(∗)**159.99**	161.18
x_2	(∗)**160.89**	161.93	160.36
x_3	(∗)163.34	(∗)161.44	(∗)159.49
x_4	(∗)163.43	(∗)162.65	(∗)159.40
x_5	(∗)165.44	(∗)163.66	(∗)160.99
x_6	(∗)162.49	(∗)161.42	(∗)160.42
x_7	(∗)163.63	(∗)162.91	(∗)161.03
x_8	(∗)164.33	(∗)162.80	(∗)160.67
x_9	(∗)165.14	(∗)163.05	(∗)161.56
x_{10}	(∗)167.23	(∗)166.92	(∗)165.00

(a) Initial model: full model

Step	(1)	(2)	(3)	(4)	(5)	(6)
Previous model: γ	–	0000000100	0000000101	0010000101	0010100101	0010101101
Action	Initialise	$+x_{10}$	$+x_3$	$+x_5$	$+x_7$	$+x_9$
Current model DIC	179.28	169.16	162.65	160.79	160.46	159.13
γ	0000000100	0000000101	0010000101	0010100101	0010101101	0010101111
x_1	181.38	170.38	163.83	161.70	160.68	160.62
x_2	180.79	167.80	162.80	161.15	162.59	159.84
x_3	173.08	**162.67**	(∗)169.03	(∗)167.46	(∗)166.86	(∗)160.96
x_4	180.66	169.87	163.16	161.37	162.42	159.22
x_5	174.33	167.37	**160.67**	(∗)162.36	(∗)161.65	(∗)161.84
x_6	177.14	168.39	163.67	161.06	160.22	159.19
x_7	178.19	167.60	161.73	**160.51**	(∗)160.79	(∗)161.88
x_8	(∗)183.79	(∗)174.58	(∗)168.44	(∗)165.10	(∗)163.57	(∗)160.42
x_9	178.61	169.94	164.67	162.05	**159.15**	(∗)160.35
x_{10}	**168.92**	(∗)179.22	(∗)173.20	(∗)169.47	(∗)170.48	(∗)164.96

(b) Initial model: model containing covariate gender

Key: (+) = add a covariate to the current model, (−) = remove a covariate from a current model, and (∗) = covariate present in the current model.

Table 6.3: Tabulated summary of stepwise procedure for grouped symptoms model, using different starting points. Bold values refer to the lowest DIC values in each step.

With these three starting models, we obtain two different best predictive models (Table 6.4). This might be because stepwise regression only adds or drops one variable

at a time. Two or more variables might have joint significant effect on the model although individually they do not, and the stepwise procedure cannot detect this.

Starting model	Covariate in final model	DIC value
Null	BMI (x_3), HBA (x_5), AWAR (x_7), GEN (x_8), TYPE (x_9), RET (x_{10})	159.13
Full	BMI (x_3), CPEP (x_4), HBA (x_5),ACE (x_6), AWAR (x_7), GEN (x_8), TYPE (x_9), RET (x_{10})	158.95
With GEN	BMI (x_3), HBA (x_5), AWAR (x_7), GEN (x_8), TYPE (x_9), RET (x_{10})	159.13

Table 6.4: Summary of the selected models from stepwise selection with different starting points for grouped symptoms model. Covariates in blue highlight the covariates which are common in the three models.

6.1.1 Stepwise selection: model with interactions between covariates

In Section 5.1.1, we discussed the effect of interaction between covariates on consistency. To decide upon which interactions are important to include in a consistency predictive model, we apply the stepwise procedure. The interactions considered for the variable selection are as given in (5.1.8) and the estimated missing values when we include those interactions are provided in Table 6.5. We initiate the procedure using the null model as the starting point. The procedure fitted 91 models within nine cycles before it was terminated. The final model contains four main predictors and six interactions between the covariates with DIC value = 153.74. Covariates gender (x_8), retinopathy (x_{10}), BMI (x_3), and awareness of hypoglycaemia (x_7) are selected as prominent variables in predicting individuals' consistency. Two interactions among them are also selected as important predictors, i.e. interaction between gender with awareness of hypoglycaemia, and BMI. The link function for the final model is as

follows:

$$\begin{aligned}\log\{E(\sigma_i^{-2})\} =& b_0 + b_3 \times BMI_i + b_7 \times AWAR_i + b_8 \times GEN_i + b_{10} \times RET_i \\ &+ b_{2\times4} \times (DUR \cdot CPEP)_i + b_{2\times10} \times (DUR \cdot RET)_i \\ &+ b_{3\times8} \times (BMI \cdot GEN)_i + b_{4\times9} \times (CPEP \cdot TYPE)_i \\ &+ b_{7\times8} \times (AWAR \cdot GEN)_i + b_{8\times10} \times (GEN \cdot RET)_i \\ i =& 1, 2, ..., 66\end{aligned}$$

and the corresponding posterior estimates for all coefficients are

$$b_0 = -0.1043 \text{ with } 95\% \text{ CI } (-0.6416, 0.4877),$$

$$b_8 = 0.9632 \text{ with } 95\% \text{ CI } (0.2316, 1.7690),$$

$$b_3 = 0.3372 \text{ with } 95\% \text{ CI } (-0.0405, 0.7217),$$

$$b_7 = 0.2148 \text{ with } 95\% \text{ CI } (0.0554, 0.3693),$$

$$b_{10,1} = 0.3185 \text{ with } 95\% \text{ CI } (0.0021, 0.6106),$$

$$b_{10,2} = -0.1718 \text{ with } 95\% \text{ CI } (-0.5013, 0.1770),$$

$$b_{10,3} = -0.1467 \text{ with } 95\% \text{ CI } (-0.5522, 0.2947),$$

$$b_{7\times8} = -0.6251 \text{ with } 95\% \text{ CI } (-0.8737, -0.3740),$$

$$b_{3\times8} = -0.5909 \text{ with } 95\% \text{ CI } (-1.0370, -0.1250),$$

$$b_{2\times4} = -0.3918 \text{ with } 95\% \text{ CI } (-0.6571, -0.0838),$$

$$b_{2\times10} = 0.1937 \text{ with } 95\% \text{ CI } (-0.1061, 0.5122).$$

When starting from the full model, the procedure was terminated after 13 cycles, fitting 131 models. Predictors age (x_1) and Hba (x_5) were deleted from the model

Covariate	Mean	SD
$CPEP_{4034}$	-0.1865	0.943
HBA_{1028}	0.1288	1.020
HBA_{1039}	0.0685	0.993
HBA_{2021}	-0.0297	1.021
HBA_{3050}	-0.0047	0.959
HBA_{4023}	-0.0205	0.967
HBA_{6019}	0.0036	1.011

Table 6.5: Posterior estimates of missing values when interactions between covariates are included.

in the first two cycles. Then, in the following cycles interaction between age and awareness ($x_{2\times7}$), duration of diabetes (x_2), C-peptide (x_4) and interaction between C-peptide and type of diabetes ($x_{4\times9}$) were also dropped from the model. The final model obtained here contains three main factors and four interactions terms, i.e.

$$\begin{aligned}\log\{E(\sigma_i^{-2})\} =& b_0 + b_3 \times BMI_i + b_8 \times GEN_i + b_{10} \times RET_i \\ &+ b_{2\times8} \times (DUR \cdot GEN)_i + b_{2\times4} \times (DUR \cdot CPEP)_i \\ &+ b_{3\times8} \times (BMI \cdot GEN)_i + b_{7\times8} \times (AWAR \cdot GEN)_i \\ i =& 1, 2, ..., 66\end{aligned}$$

with DIC value = 141.92. The posterior standardised intercept term, $b_0 = 0.1931$ with 95% CI (-0.0578,0.4735) and the posterior estimates of b_p:

$$b_8 = 1.5720 \text{ with } 95\% \text{ CI } (0.7716, 2.4520),$$

$$b_3 = 0.2652 \text{ with } 95\% \text{ CI } (-0.0910, 0.5999),$$

$$b_{10,1} = 0.6124 \text{ with } 95\% \text{ CI } (0.2580, 0.9594),$$

$$b_{10,2} = -0.4923 \text{ with } 95\% \text{ CI } (-1.0000, -0.0176),$$

$$b_{10,3} = -0.1201 \text{ with } 95\% \text{ CI } (-0.6582, 0.5418),$$

$$b_{2\times8} = -1.0800 \text{ with } 95\% \text{ CI } (-1.7240, -0.4918),$$

$$b_{7\times8} = -0.3580 \text{ with } 95\% \text{ CI } (-0.5817, -0.1192),$$

$$b_{3\times8} = -0.8622 \text{ with } 95\% \text{ CI } (-1.2860, -0.4172),$$

$$b_{2\times4} = -0.3883 \text{ with } 95\% \text{ CI } (-0.6194, -0.1486).$$

The summary tables of the stepwise procedure is not shown here but Table 6.6 summarises the final models with their DIC values obtained when we include interaction between covariates. Starting from full model gives a simpler predictor model as compared to the model obtained when starting with null model. As indicated by the blue covariates, the simpler model is almost like a subset from the other model except for interaction between duration and gender.

Starting model	Covariate in final model	DIC value
Null	BMI (x_3), AWAR (x_7), GEN (x_8), RET (x_{10}) DUR·CPEP ($x_{2\times4}$), DUR·RET ($x_{2\times10}$), CPEP·TYPE ($x_{4\times9}$), BMI·GEN ($x_{3\times8}$), AWAR·GEN ($x_{7\times8}$), GEN·RET ($x_{8\times10}$)	153.74
Full	BMI (x_3), GEN (x_8), RET (x_{10}), DUR·GEN ($x_{2\times8}$), DUR·CPEP ($x_{2\times4}$), BMI·GEN ($x_{3\times8}$), AWAR·GEN ($x_{7\times8}$)	141.92

Table 6.6: Summary of the selected models from stepwise selection with different starting points for grouped symptoms model with interaction between covariates. Covariates in blue highlight the covariates which are common in the two models.

6.1.2 Stepwise selection: hierarchical model

Previously, in Section 5.2, we constructed a hierarchical model where we can both estimate individual consistency and effect of covariates. This section will present the results from variable selection for this model using stepwise methodology. As mentioned earlier, for variable selection we need to substitute the missing values with their posterior estimates. For this model, the posterior estimates are given in Table 6.7. As before, we first started the procedure with a null model and then adding or dropping a predictor with the lowest DIC value in each cycle which we run for 10,000 iterations. Table 6.8 (a) gives a summary of the procedure. The final model is selected after six cycles with DIC value = 96.66. In the first cycle, we add variable type of diabetes (x_9), followed by retinopathy (x_{10}) in the second cycle. Then, gender (x_8) and C-peptide (x_4) are added and the final predictor to enter the model is the Ace (x_6). Therefore, the best predictive model given by stepwise regression is

$$\begin{aligned}\log\{E(\sigma_i^{-2})\} =& b_0 + b_4 \times CPEP_i + b_6 \times ACE_i + b_8 \times GEN_i + b_9 \times TYPE_i \\ &+ b_{10} \times RET_i \\ i =& 1, 2, ..., 66\end{aligned}$$

where the posterior estimates of the coefficients are:

$$
\begin{aligned}
b_0 &= 0.5016 \text{ with } 95\% \text{ CI } (0.1194, 0.8752),\\
b_4 &= -0.1715 \text{ with } 95\% \text{ CI } (-0.1715, 0.0064),\\
b_6 &= -0.1400 \text{ with } 95\% \text{ CI } (-0.3021, 0.0339),\\
b_8 &= -0.3555 \text{ with } 95\% \text{ CI } (-0.6582, -0.0391),\\
b_9 &= 0.5528 \text{ with } 95\% \text{ CI } (0.2048, 0.9904),\\
b_{10,1} &= 0.2847 \text{ with } 95\% \text{ CI } (0.07123, 0.4936),\\
b_{10,2} &= -0.1668 \text{ with } 95\% \text{ CI } (-0.4127, 0.0796),\\
b_{10,3} &= -0.1179 \text{ with } 95\% \text{ CI } (-0.4086, 0.1873). \qquad (6.1.1)
\end{aligned}
$$

Another option we have is to start the stepwise regression using all available covariates and then removing or adding one by one in each cycle depending on their DIC values. Table 6.8 (b) summarises this whole procedure which also took six cycles. Covariates age (x_1) and awareness (x_7) are eliminated from the model during the first and second cycle respectively. Duration of diabetes (x_2), Hba (x_5), and BMI (x_3) are also not very good predictors since they are eliminated by the stepwise procedure. In the end, the best predictive model only contains covariates gender (x_8), type of diabetes (x_9), retinopathy (x_{10}), C-peptide (x_4), and Ace (x_6) with the posterior estimates of the coefficients are as given in (6.1.1). This tells us that with the hierarchical model, the two starting points we used result in the same predictive model. Note that from Section 5.2, there is no significant covariate affecting consistency. Hence, the option to start from the model with significant covariate indicates that the null model is a good starting point.

Covariate	Mean	SD
$CPEP_{4034}$	-0.1382	1.013
HBA_{1028}	-0.0208	1.016
HBA_{1039}	-0.0740	0.982
HBA_{2021}	-0.0098	0.963
HBA_{3050}	-0.0841	1.033
HBA_{4023}	-0.1356	1.013
HBA_{6019}	-0.1274	1.028

Table 6.7: Posterior estimates of missing values in hierarchical model.

Step	(1)	(2)	(3)	(4)	(5)	(6)
Previous model: γ	$-$	0000000000	0000000010	0000000011	0000000111	0001000111
Action	Initialise	$+x_9$	$+x_{10}$	$+x_8$	$+x_4$	$+x_6$
Current model DIC	112.20	104.11	101.51	98.44	97.72	96.66
γ	0000000000	0000000010	0000000011	0000000111	0001000111	0001010111
x_1	113.29	106.13	102.98	101.05	99.23	98.73
x_2	109.11	104.31	102.93	99.58	99.41	98.19
x_3	111.22	105.40	101.56	98.63	98.35	98.15
x_4	114.14	104.83	101.03	**97.45**	(∗)98.34	(∗)98.40
x_5	113.93	105.68	103.30	101.02	99.38	98.64
x_6	107.74	103.25	99.95	97.76	**97.22**	(∗)97.24
x_7	113.98	105.43	102.22	100.57	100.39	98.22
x_8	109.68	102.84	**98.46**	(∗)101.01	(∗)100.91	(∗)98.47
x_9	**104.01**	(∗)112.16	(∗)106.83	(∗)103.25	(∗)105.45	(∗)102.99
x_{10}	106.86	**100.92**	(∗)104.27	(∗)102.35	(∗)100.78	(∗)100.95

(a) Initial model: null model

Step	(1)	(2)	(3)	(4)	(5)	(6)
Previous model: γ	$-$	1111111111	0111111111	0111110111	0011110111	0011010111
Action	Initialise	$-x_1$	$-x_7$	$-x_2$	$-x_5$	$-x_3$
Current model DIC	105.62	103.03	101.89	99.60	97.49	96.66
γ	1111111111	0111111111	0111110111	0011110111	0011010111	0001010111
x_1	(∗)**103.69**	105.63	103.38	102.18	99.22	98.73
x_2	(∗)104.10	(∗)102.22	(∗)**98.99**	101.74	99.39	98.19
x_3	(∗)104.78	(∗)103.15	(∗)100.03	(∗)98.63	(∗)**96.82**	98.15
x_4	(∗)106.83	(∗)104.84	(∗)102.44	(∗)100.91	(∗)98.71	(∗)98.40
x_5	(∗)104.51	(∗)101.74	(∗)99.24	(∗)**98.10**	99.65	98.64
x_6	(∗)106.00	(∗)104.47	(∗)101.77	(∗)99.93	(∗)97.31	(∗)97.24
x_7	(∗)104.00	(∗)**100.69**	103.90	101.82	99.17	98.22
x_8	(∗)107.20	(∗)105.12	(∗)103.52	(∗)101.96	(∗)100.42	(∗)98.47
x_9	(∗)108.34	(∗)106.43	(∗)104.98	(∗)103.23	(∗)101.60	(∗)102.99
x_{10}	(∗)105.28	(∗)103.72	(∗)102.23	(∗)103.56	(∗)102.71	(∗)100.95

(b) Initial model: full model

Key: (+) = add a covariate to the current model, (−) = remove a covariate from a current model, and (∗) = covariate present in the current model.

Table 6.8: Tabulated summary of stepwise procedure for hierarchical model with grouped symptoms, using different starting points. Bold values refer to the lowest DIC values in each step.

6.2 Gibbs Variable Selection

This section discusses another method we use to select patient-specific covariates that best describe the predictive model for consistency of symptom reporting for individual patients. The method used here is Gibbs Variable selection, which was introduced by Dellaportas et al. (2002). We consider three different sets of prior distributions for comparison purposes. This section focuses only on predictive models arising from the grouped symptoms model.

As before, covariates age (x_1), duration of diabetes (x_2), BMI (x_3), C-peptide (x_4), haemoglobin A1c (Hba) (x_5), serum angiotensin converting enzyme (Ace) (x_6), awareness of hypoglycaemia(x_7), gender (x_8), type of diabetes (x_9) and retinopathy (x_{10}) are considered for inclusion in our predictive model. With $p = 10$ patient-specific covariates in the model, we have a model space containing all potential models, $\mathcal{M} = \{\mathcal{M}_1, ..., \mathcal{M}_{2^p}\}$. The 1024 possible models include the model with no covariates or what we call a null model.

We introduce a $p \times 1$ binary indicator vector, $\boldsymbol{\gamma}$ into the linear predictor where p is the number of covariates under consideration. Thus we have,

$$\log(m_i) = b_0 + \sum_{j=1}^{9} \gamma_j b_j X_j + \gamma_{10} X_{ret}.$$

The indicator variable, $\boldsymbol{\gamma}$, will identify which covariate will be included in t^{th} model, namely $\mathcal{M}_t$. The likelihood of $\mathcal{M}_t$ is given by,

$$f(\mathbf{D}|\mathcal{M}_t) = \int f(\mathbf{D}|\boldsymbol{b_t}, \mathcal{M}_t)\pi(\boldsymbol{b_t}|\mathcal{M}_t)db_t,$$

where $\boldsymbol{b_t}$ denotes the parameter vector in model $\mathcal{M}_t$. $f(\mathbf{D}|\boldsymbol{b_t}, \mathcal{M}_t)$ is the likelihood function under model $\mathcal{M}_t$, conditional on $\boldsymbol{b_t}$ and $\pi(\boldsymbol{b_t}|\mathcal{M}_t)$ is the prior density of the parameter vector $\boldsymbol{b_t}$ (Kass and Raftery, 1995). The prior corresponding to model $\mathcal{M}_t$

is $\pi(\mathcal{M}_t)$ and the posterior probability of $\mathcal{M}_t$ is calculated using Bayes Theorem:

$$p(\mathcal{M}_t|D) = \frac{f(\mathbf{D}|\mathcal{M}_t)\pi(\mathcal{M}_t)}{\sum_{t=1}^{1024} f(\mathbf{D}|\mathcal{M}_t)\pi(\mathcal{M}_t)}. \tag{6.2.1}$$

To compare any two models we use posterior odds (Jeffreys, 1946) which can be evaluated using ratios of the posterior probabilities from (6.2.1), to get the posterior odds of the models, PO_{01}

$$PO_{01} = \frac{p(\mathcal{M}_0|\mathbf{D})}{p(\mathcal{M}_1|\mathbf{D})} = \frac{f(\mathbf{D}|\mathcal{M}_0)}{f(\mathbf{D}|\mathcal{M}_1)} \times \frac{\pi(\mathcal{M}_0)}{\pi(\mathcal{M}_1)}, \tag{6.2.2}$$

where $\frac{f(\mathbf{D}|\mathcal{M}_0)}{f(\mathbf{D}|\mathcal{M}_1)}$ is the Bayes factor for $\mathcal{M}_0$ against $\mathcal{M}_1$ and $\frac{\pi(\mathcal{M}_0)}{\pi(\mathcal{M}_1)}$ is the prior odds ratio. (6.2.2) can be written as, posterior odds = Bayes factor $\times$ prior odds. $PO_{01} > 1$ will implies that the data favour $\mathcal{M}_0$ and vice versa.

We can also monitor the variable inclusion probability, $p(\gamma_j|\mathbf{D})$, which is the probability of variable j be included in all visited models throughout a simulation. Therefore we have:

$$p(\gamma_j|\mathbf{D}) = \sum_{\gamma_{\setminus j} \in \{0,1\}^{p-1}} p(\gamma_j, \gamma_{\setminus j}|\mathbf{D})$$

Besides our main interest to estimate the maximum a posteriori (MAP) model, alternatively we can trace a median probability (MP) model which refers to the model which includes all variable with $p(\gamma_j|\mathbf{D}) > 0.5$ (Ntzoufras, 2009).

Choosing prior distributions is crucial in Bayesian model selection as the posterior model probabilities depend very much on the prior. This situation relates to the Lindley Paradox, where a more parsimonious model is suggested with higher prior variance. We use $\gamma \sim Bernoulli(0.5)$ to represent lack of information on which covariates are important whereas the prior for b_j is:

$$b_j \sim N(\mu_j, \sigma_j^2). \tag{6.2.3}$$

The independent prior, Zellner's g-prior, and empirical prior are the three different sets of parameters for prior distribution of b_j explored in this study. Detailed discussions are provided in next sections.

Independent priors

The first prior to consider is the independent prior originally used by Dellaportas et al. (2002). Parameters μ_j and σ_j^2 in Equation (6.2.3) are defined as follows:

$$\mu_j = (1 - \gamma_j)\bar{\mu}_{b_j}$$

$$\sigma_j^2 = \gamma_j c^2 + (1 - \gamma_j)\bar{\sigma}_{b_j}^2$$

where $\bar{\mu}_{b_j}$ and $\bar{\sigma}_{b_j}^2$ are the posterior mean and variance of b_j respectively, obtained from a pilot run of the full model while c^2 is a constant. Multiple values for constant c^2 are used for comparison and to study the effect of the Lindley-Bartlett paradox. We use $c^2 = n$, 1000 and 10,000. The sample size, $n = 66$, is the number of patients.

Table 6.9 (a) provides the mean and standard deviation (SD) of the posterior inclusion probabilities of the γ_j variables, corresponding to X_j covariates. These are obtained with an MCMC run of 10,000 iterations after a 5000 burn-in period. With $c^2 = n$, gender (γ_8) is the only covariate with high inclusion probability, 0.8098. As c^2 increases to 10,000 gender's probability of inclusion decreases to just slightly over 0.5, that is 0.5048. The other nine covariates have probabilities less than 0.5. We notice that the probability of inclusion of all covariates decreases as the prior variance increases.

Tables 6.9 (b) - (d) present the ten predictive models with the highest estimated posterior model probabilities and posterior odds compared to the top ranked model. The model with the highest estimated posterior probability is called the maximum a posteriori (MAP) model. When factor $c^2 = n$, the MAP model is $\mathcal{M}_{129}$ and contains covariate gender (γ_8). We can also look at the median probability (MP) model, which

is the model that includes all covariates with posterior inclusion probability at least 0.5. In this case, our MP model is also $\mathcal{M}_{129}$. Kass and Raftery (1995) state that if the Bayes factor between models is less than 3, the differences between them can be ignored. $\mathcal{M}_{145}$ which contain covariates Hba (γ_5) and gender (γ_8) has probability 0.1803 and posterior odd 1.31. Thus can we say that these two models are equally good predictive models. $\mathcal{M}_{129}$ is also suggested as the MP and MAP model if we use c^2= 1000 and 10,000.

Zellner's g - prior

The next prior we consider is the Zellner's g-prior (Zellner, 1986). With this setting, parameter $\boldsymbol{b}$ takes a multivariate normal prior distribution with mean $\boldsymbol{\mu_b}$ and precision $\boldsymbol{S^{-1}}$.

$$\boldsymbol{b} \sim NVM(\boldsymbol{\mu_b}, \boldsymbol{S}^{-1})$$

$$\text{where } \boldsymbol{\mu_b} = \mu_0, ..., \mu_p$$

$$\text{with } \mu_j = (1 - \gamma_j)\bar{\mu}_{b_j} \text{ for } j = 1, ..., p.$$

Define $\boldsymbol{S^{-1}} = c^2[\mathbf{X}^\intercal\mathbf{X}]^{-1}\sigma^2$, so each element of matrix $\boldsymbol{S}$ is:

$$S_{j,k} = \frac{\gamma_j\gamma_k}{c^2\sigma^2}[\mathbf{X}^\intercal\mathbf{X}]_{jk} + (1 - \gamma_j\gamma_k)I(j = k)\bar{S}^2_{b_j}$$

for $j, k = 1, 2, ..., p$. $\mathbf{X}$ is the $n \times (p + 1)$ standardised design matrix whereas c^2 is a constant. As described previously, $\bar{\mu}_{b_j}$ and $\bar{S}^2_{b_j}$ take values of the estimated mean and standard deviation of b_j from a pilot run of the full model and $\gamma_0 = 1$.

As previously, we use three different c^2 values; $c^2 = n, 1000$, and 10,000. The full results of the posterior variable inclusion probabilities and the posterior model probabilities are given in Table 6.10. The probability of inclusion of covariate gender (γ_8) is always greater than 0.5 even with a large prior variance. However, we can see the probability of inclusion for every covariate decreases as the prior variance gets higher.

This trend can be clearly seen in the probability of inclusion for retinopathy (γ_{10}) where it has declined from 0.9419 (with $c^2 = 66$) to 0.3116 (with $c^2 = 10,000$).

Referring to Table 6.10 (b), the MAP model is $\mathcal{M}_{661}$ when $c^2 = 66$. This model includes covariates BMI (γ_3), Hba (γ_5), gender (γ_8), and retinopathy (γ_{10}) and has probability of 0.0326. We can see that all ten models presented here have small posterior odds suggesting high model uncertainty. The MP model is $\mathcal{M}_{721}$ (not shown here), with the posterior odds when comparing it to the MAP model being 3.34. As the prior variance gets higher, i.e. by using $c^2 = 1000$ and 10,000, more parsimonious models are suggested. $\mathcal{M}_{641}$, $\mathcal{M}_{645}$, and $\mathcal{M}_{145}$ are favoured in these cases. With $c^2 = 1000$, the MAP and MP models contain covariates gender and retinopathy, i.e. $\mathcal{M}_{641}$, whereas with $c^2 = 10,000$ the MAP and MP selected models only have gender as predictor, i.e. $\mathcal{M}_{129}$.

Empirical prior

The third prior we consider is the empirical prior (Ntzoufras, 2011). We substitute the parameters in Equation (6.2.3) as follows:

$$\mu_j = \bar{\mu}_{b_j}$$

$$\sigma_j^2 = \{\gamma_j n + (1 - \gamma_j)\}\, \bar{\sigma}_{b_j}^2$$

where j = 1,2,...,p. $\bar{\mu}_{b_j}$ is the posterior b_j estimates obtained from a MCMC pilot run of the full model with 10,000 iterations. $\bar{\sigma}_{b_j}^2$ is the posterior variance of b_j from the same run. This prior setting is exactly the same with the independent prior when covariate j is not in the model (i.e. when $\gamma_j = 0$), but is different when $\gamma_j = 1$.

Based on the variable inclusion probabilities in Table 6.11 (a), variables gender (γ_8) and retinopathy (γ_{10}) have high probabilities, 0.9786 and 0.9419 respectively, making these two covariates almost always included in all visited models. Other covariates with probability of inclusion greater than 0.5 are type of diabetes (γ_9) and Hba (γ_5).

The MAP model suggested by this prior is $\mathcal{M}_{915}$ (Table 6.11 (b)). Besides gender and retinopathy, this model contains variables duration of diabetes (γ_2), Hba (γ_5), and type of diabetes (γ_9), and has posterior probability 0.0695. Since covariate duration has probability of inclusion less than 0.5, it is not included in the MP model which is $\mathcal{M}_{913}$. This model has probability 0.0134 and the posterior odd of $\mathcal{M}_{915}$ to $\mathcal{M}_{913}$ is 5.21 implying that with this prior setting, the MAP model is five times better than the MP model. There are four other models that have posterior odd less than 3 if compared to the MAP model. With posterior odd 1.40, we can say that $\mathcal{M}_{645}$ with probability 0.0495 is almost as good as the MAP model. $\mathcal{M}_{661}$, $\mathcal{M}_{647}$, and $\mathcal{M}_{899}$ have probabilities less than the MAP models but their posterior odd are still below 3 indicating that these models are not significantly different from each other.

Note that gender is included in all selected models. Thus, we can say that this covariate is prominent in predicting the consistency of individual patients. Together with this, retinopathy and Hba also have high probability of inclusion and are selected as important predictors by all three prior settings. Although different prior settings come out with different 'best' models (MAP), they all have a small number of similar models selected such as $\mathcal{M}_{661}$, $\mathcal{M}_{641}$, $\mathcal{M}_{645}$, and $\mathcal{M}_{145}$. Again, this is an indication of model uncertainty.

6.2.1 Analysis with 59 patients

During the earlier phase of our work, we were only provided with data on 59 patients before we received additional data on seven subjects. Therefore, we also have results for Gibbs variable selection using only 59 patients. We present here the results when using Zellner's g-prior. Table 6.12 gives the inclusion probabilities for each covariate and the posterior model probabilities.

Observing the inclusion probabilities of all covariates, we notice obvious differences in inclusion probabilities of covariates Hba (γ_5) and retinopathy (γ_{10}). Hba does not have very high inclusion probability if we include the seven additional patients (see Table 6.10), except when using $c^2 = n$ where it has probability of inclusion above 0.5.

On the other hand, when the seven patients are excluded (Table 6.12) the probability is high at any level of c^2. As a result, the 'best' models selected by this procedure always include Hba. As for retinopathy, this covariate has high probability with the seven patients in the analysis but decreases if we take them out. Therefore, the best models do not contain this covariate, whereas in the previous analysis it is included in almost each of the selected models.

When $c^2 = n$ the MAP model in Table 6.12 (b) is $\mathcal{M}_{209}$, which contains covariates Hba (γ_5), awareness of hypoglycaemia (γ_7), and gender (γ_8), and has posterior model probability of 0.1102. This is similar to the outcome if we use $c^2 = 1000$ where the MAP model is also $\mathcal{M}_{209}$ but with higher probability, 0.3409.

Generally, models selected when using higher prior variance are much simpler than models selected when $c^2 = 59$. If $c^2 = 10{,}000$, we obtain three models with posterior odds ≤ 3 which are $\mathcal{M}_{145}$, $\mathcal{M}_{209}$, and $\mathcal{M}_{193}$. $\mathcal{M}_{145}$ is the MAP and MP model and contain covariate Hba (γ_5) and gender (γ_8).

Based on this finding, results here are different from when these seven patients are included in the analysis. Covariates Hba (γ_5) and retinopathy (γ_{10}) are highly affected with the inclusion or exclusion of the seven patients. Hence, it seems like this method is sensitive to changes in data. This ambiguous result may be due to small data size (i.e. 59 and 66 patients) rather than to the variable having or not having appreciable predictive value.

(a) Inclusion probabilities for various prior variances.

	$c^2 = n$		$c^2 = 1000$		$c^2 = 10{,}000$	
Parameter	$p(\gamma\|\mathbf{D})$	SD	$p(\gamma\|\mathbf{D})$	SD	$p(\gamma\|\mathbf{D})$	SD
γ_1	0.0181	0.13	0.0039	0.06	0.0012	0.03
γ_2	0.0467	0.21	0.0082	0.09	0.0011	0.03
γ_3	0.1158	0.32	0.0446	0.21	0.0235	0.15
γ_4	0.0363	0.19	0.0082	0.09	0.0021	0.05
γ_5	0.3843	0.49	0.1865	0.39	0.0931	0.29
γ_6	0.1192	0.32	0.0495	0.22	0.0213	0.14
γ_7	0.1648	0.37	0.0316	0.17	0.0133	0.11
γ_8	0.8098	0.39	0.7214	0.45	0.5048	0.50
γ_9	0.1597	0.37	0.0285	0.17	0.0070	0.08
γ_{10}	0.1709	0.38	0.0184	0.13	0.0013	0.04

(b) Posterior model probabilities - $c^2 = n$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{129}/\mathcal{M}_t)$
1	$\mathcal{M}_{129}$	x_8	0.2353	1.00
2	$\mathcal{M}_{145}$	$x_5 + x_8$	0.1803	1.31
3	$\mathcal{M}_{641}$	$x_8 + x_{10}$	0.0733	3.21
4	$\mathcal{M}_{133}$	$x_3 + x_8$	0.0373	6.31
5	$\mathcal{M}_{321}$	$x_7 + x_9$	0.0298	7.90
6	$\mathcal{M}_{17}$	x_5	0.0256	9.19
7	$\mathcal{M}_{401}$	$x_5 + x_8 + x_9$	0.0242	9.72
8	$\mathcal{M}_{177}$	$x_5 + x_6 + x_8$	0.0207	11.37
9	$\mathcal{M}_{645}$	$x_3 + x_8 + x_{10}$	0.0205	11.48
10	$\mathcal{M}_{161}$	$x_6 + x_8$	0.0196	12.01

(c) Posterior model probabilities - $c^2 = 1000$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{129}/\mathcal{M}_t)$
1	$\mathcal{M}_{129}$	x_8	0.5330	1.00
2	$\mathcal{M}_1$	constant	0.1415	3.77
3	$\mathcal{M}_{145}$	$x_5 + x_8$	0.0988	5.39
4	$\mathcal{M}_{17}$	x_5	0.0587	9.08
5	$\mathcal{M}_{133}$	$x_3 + x_8$	0.0233	22.88
6	$\mathcal{M}_{33}$	x_6	0.0181	29.45
7	$\mathcal{M}_5$	x_3	0.0140	38.07
8	$\mathcal{M}_{161}$	$x_6 + x_8$	0.0123	43.33
9	$\mathcal{M}_{641}$	$x_8 + x_{10}$	0.0094	56.70
10	$\mathcal{M}_{193}$	$x_7 + x_8$	0.0088	60.57

(d) Posterior model probabilities - $c^2 = 10{,}000$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{129}/\mathcal{M}_t)$
1	$\mathcal{M}_{129}$	x_8	0.4615	1.00
2	$\mathcal{M}_1$	constant	0.3839	1.20
3	$\mathcal{M}_{17}$	x_5	0.0613	7.53
4	$\mathcal{M}_{145}$	$x_5 + x_8$	0.0256	18.03
5	$\mathcal{M}_5$	x_3	0.0150	30.77
6	$\mathcal{M}_{33}$	x_6	0.0142	32.50
7	$\mathcal{M}_{65}$	x_7	0.0082	56.28
8	$\mathcal{M}_{133}$	$x_3 + x_8$	0.0073	63.22
9	$\mathcal{M}_{49}$	$x_5 + x_6$	0.0026	177.50
10	$\mathcal{M}_{161}$	$x_6 + x_8$	0.0024	192.29

Table 6.9: Parameter inclusion probabilities and model probabilities with independent normal prior.

(a) Inclusion probabilities for various prior variances.

	$c^2 = n$		$c^2 = 1000$		$c^2 = 10{,}000$	
Parameter	$p(\gamma\|\mathbf{D})$	SD	$p(\gamma\|\mathbf{D})$	SD	$p(\gamma\|\mathbf{D})$	SD
γ_1	0.1659	0.37	0.0484	0.21	0.0165	0.13
γ_2	0.2966	0.46	0.1951	0.40	0.0567	0.23
γ_3	0.4672	0.50	0.2456	0.43	0.1206	0.33
γ_4	0.2378	0.43	0.0770	0.27	0.0177	0.13
γ_5	0.6120	0.49	0.3610	0.48	0.2986	0.46
γ_6	0.4148	0.49	0.2592	0.44	0.1464	0.35
γ_7	0.6198	0.49	0.4196	0.49	0.1826	0.39
γ_8	0.8659	0.34	0.7498	0.43	0.7356	0.44
γ_9	0.4124	0.49	0.2182	0.41	0.0672	0.25
γ_{10}	0.9419	0.23	0.7459	0.44	0.3116	0.46

(b) Posterior model probabilities - $c^2 = n$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{661}/\mathcal{M}_t)$
1	$\mathcal{M}_{661}$	$x_3 + x_5 + x_8 + x_{10}$	0.0326	1.00
2	$\mathcal{M}_{725}$	$x_3 + x_5 + x_7 + x_8 + x_{10}$	0.0265	1.23
3	$\mathcal{M}_{709}$	$x_3 + x_7 + x_8 + x_{10}$	0.0237	1.37
4	$\mathcal{M}_{977}$	$x_5 + x_7 + x_8 + x_9 + x_{10}$	0.0237	1.37
5	$\mathcal{M}_{753}$	$x_5 + x_6 + x_7 + x_8 + x_{10}$	0.0196	1.66
6	$\mathcal{M}_{663}$	$x_2 + x_3 + x_5 + x_8 + x_{10}$	0.0192	1.70
7	$\mathcal{M}_{737}$	$x_6 + x_7 + x_8 + x_{10}$	0.0185	1.76
8	$\mathcal{M}_{1009}$	$x_5 + x_6 + x_7 + x_8 + x9 + x10$	0.0180	1.81
9	$\mathcal{M}_{645}$	$x_3 + x_8 + x_{10}$	0.0180	1.81
10	$\mathcal{M}_{647}$	$x_2 + x_3 + x_8 + x_{10}$	0.0161	2.02

(c) Posterior model probabilities - $c^2 = 1000$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{641}/\mathcal{M}_t)$
1	$\mathcal{M}_{641}$	$x_8 + x_{10}$	0.0863	1.00
2	$\mathcal{M}_{645}$	$x_3 + x_8 + x_{10}$	0.0608	1.42
3	$\mathcal{M}_{643}$	$x_2 + x_8 + x_{10}$	0.0577	1.50
4	$\mathcal{M}_{705}$	$x_7 + x_8 + x_{10}$	0.0408	2.12
5	$\mathcal{M}_{609}$	$x_6 + x_7 + x_{10}$	0.0370	2.33
6	$\mathcal{M}_{657}$	$x_5 + x_8 + x_{10}$	0.0321	2.69
7	$\mathcal{M}_{145}$	$x_5 + x_8$	0.0294	2.94
8	$\mathcal{M}_{661}$	$x_3 + x_5 + x8 + x_{10}$	0.0232	3.72
9	$\mathcal{M}_{647}$	$x_2 + x_3 + x8 + x_{10}$	0.0214	4.03
10	$\mathcal{M}_{737}$	$x_6 + x_7 + x8 + x_{10}$	0.0198	4.36

(d) Posterior model probabilities - $c^2 = 10{,}000$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{641}/\mathcal{M}_t)$
1	$\mathcal{M}_{129}$	x_8	0.2055	0.63
2	$\mathcal{M}_{641}$	$x_8 + x_{10}$	0.1304	1.00
3	$\mathcal{M}_{145}$	$x_5 + x_8$	0.1265	1.03
4	$\mathcal{M}_{17}$	x_5	0.0401	3.25
5	$\mathcal{M}_{645}$	$x_3 + x_8 + x_{10}$	0.0335	3.89
6	$\mathcal{M}_{97}$	$x_6 + x_7$	0.0318	4.10
7	$\mathcal{M}_{133}$	$x_3 + x_8$	0.0278	4.69
8	$\mathcal{M}_1$	constant	0.0267	4.88
9	$\mathcal{M}_{643}$	$x_2 + x_8 + x_{10}$	0.0240	5.43
10	$\mathcal{M}_{193}$	$x_7 + x_8$	0.0227	5.74

Table 6.10: Parameter inclusion probabilities and model probabilities with Zellner's g-prior.

(a) Inclusion probabilities.

Parameter	$p(\gamma\|\mathbf{D})$	SD
γ_1	0.1077	0.31
γ_2	0.4783	0.50
γ_3	0.4130	0.49
γ_4	0.2446	0.43
γ_5	0.5493	0.50
γ_6	0.2482	0.43
γ_7	0.2651	0.44
γ_8	0.9786	0.14
γ_9	0.5664	0.50
γ_{10}	0.9419	0.23

(b) Posterior model probabilities.

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{915}/\mathcal{M}_t)$
1	$\mathcal{M}_{915}$	$x_2 + x_5 + x_8 + x_9 + x_{10}$	0.0695	1.00
2	$\mathcal{M}_{645}$	$x_3 + x_8 + x_{10}$	0.0495	1.40
3	$\mathcal{M}_{661}$	$x_3 + x_5 + x_8 + x_{10}$	0.0363	1.92
4	$\mathcal{M}_{647}$	$x_2 + x_3 + x_8 + x_{10}$	0.0332	2.09
5	$\mathcal{M}_{899}$	$x_2 + x_8 + x_9 + x_{10}$	0.0314	2.22
6	$\mathcal{M}_{903}$	$x_2 + x_3 + x_8 + x_9 + x_{10}$	0.0219	3.17
7	$\mathcal{M}_{919}$	$x_2 + x_3 + x_5 + x_8 + x_9 + x_{10}$	0.0216	3.22
8	$\mathcal{M}_{663}$	$x_2 + x_3 + x_5 + x_8 + x_{10}$	0.0201	3.47
9	$\mathcal{M}_{923}$	$x_2 + x_4 + x_5 + x_8 + x_9 + x_{10}$	0.0189	3.68
10	$\mathcal{M}_{977}$	$x_5 + x_7 + x_8 + x_9 + x_{10}$	0.0181	3.85

Table 6.11: Parameter inclusion probabilities and model probabilities with empirical prior.

(a) Inclusion probabilities for various prior variances.

	$c^2 = n$		$c^2 = 1000$		$c^2 = 10{,}000$	
Parameter	$p(\gamma\|\mathbf{D})$	SD	$p(\gamma\|\mathbf{D})$	SD	$p(\gamma\|\mathbf{D})$	SD
γ_1	0.1347	0.34	0.0348	0.18	0.0105	0.10
γ_2	0.2859	0.45	0.0695	0.25	0.0162	0.13
γ_3	0.3029	0.46	0.0710	0.26	0.0212	0.14
γ_4	0.1392	0.35	0.0422	0.20	0.0156	0.12
γ_5	0.9201	0.27	0.8768	0.33	0.7836	0.41
γ_6	0.1575	0.36	0.0443	0.21	0.0170	0.13
γ_7	0.7501	0.43	0.5866	0.49	0.4269	0.49
γ_8	0.9906	0.10	0.9798	0.14	0.9295	0.26
γ_9	0.1497	0.36	0.0467	0.21	0.0144	0.12
γ_{10}	0.6359	0.48	0.1211	0.33	0.0122	0.11

(b) Posterior model probabilities - $c^2 = n$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{209}/\mathcal{M}_t)$
1	$\mathcal{M}_{209}$	$x_5 + x_7 + x_8$	0.1102	1.00
2	$\mathcal{M}_{721}$	$x_5 + x_7 + x_8 + x_{10}$	0.0958	1.15
3	$\mathcal{M}_{725}$	$x_3 + x_5 + x_7 + x_8 + x_{10}$	0.0612	1.80
4	$\mathcal{M}_{659}$	$x_2 + x_5 + x_8 + x_{10}$	0.0390	2.83
5	$\mathcal{M}_{723}$	$x_2 + x_5 + x_7 + x_8 + x_{10}$	0.0368	2.99
6	$\mathcal{M}_{213}$	$x_3 + x_5 + x_7 + x_8$	0.0237	4.65
7	$\mathcal{M}_{663}$	$x_2 + x_3 + x_5 + x_8 + x_{10}$	0.0226	4.89
8	$\mathcal{M}_{241}$	$x_5 + x_6 + x_7 + x_8$	0.0224	4.92
9	$\mathcal{M}_{753}$	$x_5 + x_6 + x_7 + x_8 + x_{10}$	0.0212	5.20
10	$\mathcal{M}_{727}$	$x_2 + x_3 + x_5 + x_7 + x_8 + x_{10}$	0.0201	5.48

(c) Posterior model probabilities - $c^2 = 1000$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{209}/\mathcal{M}_t)$
1	$\mathcal{M}_{209}$	$x_5 + x_7 + x_8$	0.3409	1.00
2	$\mathcal{M}_{145}$	$x_5 + x_8$	0.2608	1.31
3	$\mathcal{M}_{193}$	$x_7 + x_8$	0.0557	6.12
4	$\mathcal{M}_{721}$	$x_5 + x_7 + x_8 + x_{10}$	0.0236	14.48
5	$\mathcal{M}_{149}$	$x_3 + x_5 + x_8$	0.0225	15.15
6	$\mathcal{M}_{659}$	$x_2 + x_5 + x_8 + x_{10}$	0.0179	19.10
7	$\mathcal{M}_{213}$	$x_3 + x_5 + x_7 + x_8$	0.0174	19.65
8	$\mathcal{M}_{465}$	$x5 + x7 + x8 + x9$	0.0166	20.54
9	$\mathcal{M}_{241}$	$x_5 + x_6 + x_7 + x_8$	0.0160	21.37
10	$\mathcal{M}_{705}$	$x_7 + x_8 + x_{10}$	0.0157	21.71

(d) Posterior model probabilities - $c^2 = 10{,}000$

Rank	Model No	Model	$f(\mathcal{M}\|\mathbf{D})$	$PO(\mathcal{M}_{209}/\mathcal{M}_t)$
1	$\mathcal{M}_{145}$	$x_5 + x_8$	0.4940	0.43
2	$\mathcal{M}_{209}$	$x_5 + x_7 + x_8$	0.2102	1.00
3	$\mathcal{M}_{193}$	$x_7 + x_8$	0.1193	1.76
4	$\mathcal{M}_{65}$	x_7	0.0475	4.43
5	$\mathcal{M}_{129}$	x_8	0.0189	11.12
6	$\mathcal{M}_{149}$	$x_3 + x_5 + x_8$	0.0128	16.49
7	$\mathcal{M}_{153}$	$x_4 + x_5 + x_8$	0.0079	26.61
8	$\mathcal{M}_{81}$	$x_5 + x_7$	0.0075	28.21
9	$\mathcal{M}_{177}$	$x_5 + x_6 + x_8$	0.0072	29.40
10	$\mathcal{M}_{401}$	$x_5 + x_8 + x_9$	0.0066	31.85

Table 6.12: Parameter inclusion probabilities and model probabilities with Zellner's g-prior for 59 patients.

6.3 Bayesian Model Averaging

Model uncertainty is a critical issue in model selection. Model selection often gives a single 'best' model, thus giving us false impression of which model best explains the data. Moreover, using different methods for model selection such as Gibbs variable selection and stepwise regression will many times lead to different final models (Weisberg, 1985).

In the previous section, the Gibbs variable selection suggests a different 'best' model when using different priors. Also, with each prior, the posterior model probability does not strongly concentrate on only one model. We have at least two models with posterior model probability odds between them ≤ 3, making it difficult to be confident about which model makes the best predictive model and to leave out other models. To deal with this model uncertainty problem we will use Bayesian model averaging (Hoeting et al., 1999). Instead of using one single model for prediction, this method averages a selected set of models. Madigan and Raftery (1994) measure the performance of model averaging using logarithmic scoring rule and conclude that averaging over all selected models gives better predictive ability than using a single model. However, one needs to remember that this method does not tell us which variables are important in the prediction model.

It will be time consuming to average all possible models in the model space, $\mathcal{M}$. Thus Bayesian model averaging sums over only non-negligible models, which are models that have posterior odds ≤ 3, as compared to the 'best' model, since they are not significantly different from each other. For each prior case, we select models $\mathcal{M}_t$ with posterior odds ≤ 3. Recall that from Section 2.5, individual patient consistency $\widetilde{c}_i$ is a function of $\widetilde{w}_i = \dfrac{1}{\sigma_i^2}$. Hence, this parameter is used to compute a model averaged predicted consistency for patient i, by mixing the inference from selected models $\mathcal{M}_t$ using their posterior probabilities as weights. Therefore, we have

$$\widetilde{w}_i^{(A)} = \sum_{k=1}^{K} \widetilde{w}_i^{(k)} P(\mathcal{M}_{t_k}|D) \tag{6.3.1}$$

$$= \widetilde{w}_i^{(1)} P(\mathcal{M}_{t_1}|D) + \widetilde{w}_i^{(2)} P(\mathcal{M}_{t_2}|D) + ... + \widetilde{w}_i^{(K)} P(\mathcal{M}_{t_k}|D) \tag{6.3.2}$$

in which k =1,...,K is the sampled models and $P(\mathcal{M}_{t_k}|D)$ is the posterior model probability for model k.

As discussed previously, the Gibbs variable selection in Section 6.2 shows evidence of model uncertainty. For instance, the outcome when using Zellner's g-prior (with $c^2 = n$) in Table 6.10 tells us that $\mathcal{M}_{661}$ is the 'best' model with posterior probability 0.0326. Nonetheless, this model only represents 3.26% of the total posterior model probability indicating there is a fair amount of model uncertainty. To tackle this uncertainty issue, we take an average of models, using Bayesian model averaging so that instead of making conclusion only on model $\mathcal{M}_{661}$, other potential models are taken into account too.

From the results in Tables 6.9- 6.11, we select models with posterior odds ≤ 3 and summarise them in Table 6.13. This table lists the selected models together with their corresponding covariates and posterior model probability, $f(\mathcal{M}|\mathbf{D})$. From the table, it appears that covariate gender (x_8) is a common variable for all the selected models. Retinopathy (x_{10}) also appears in all selected models except models selected when using independent prior.

We elaborate here an example of how we conduct the model averaging. This example mixes over potential best models obtained when using Zellner's g-prior. There are 15 models with posterior odds ≤ 3, however we limit the averaging to only 10 models because remaining models have very small probability. These models account for about 22% of the total posterior probability. Therefore, we normalise these probabilities over all selected models so that their sum equal to 1. The normalised probabilities, $f^*(\mathcal{M}|\mathbf{D})$ are given in the last column in Table 6.13. We then conduct MCMC simulation of 20,000 iterations of each selected model $\mathcal{M}_k$ and monitor the value of $\widetilde{w}_i^{(k)}$.

	Covariates									
Selected model	x_2	x_3	x_5	x_6	x_7	x_8	x_9	x_{10}	$f(\mathcal{M}\|\mathbf{D})$	$f^*(\mathcal{M}\|\mathbf{D})$
Independent prior										
$\mathcal{M}_{129}$						•			0.2353	0.5662
$\mathcal{M}_{145}$			•			•			0.1803	0.4338
Zellner's g-prior										
$\mathcal{M}_{661}$		•	•			•		•	0.0326	0.1510
$\mathcal{M}_{725}$		•	•		•	•		•	0.0265	0.1227
$\mathcal{M}_{709}$		•			•	•		•	0.0237	0.1098
$\mathcal{M}_{977}$			•		•	•	•	•	0.0237	0.1098
$\mathcal{M}_{753}$			•	•	•	•		•	0.0196	0.0908
$\mathcal{M}_{663}$	•	•	•			•		•	0.0192	0.0889
$\mathcal{M}_{737}$				•	•	•		•	0.0185	0.0857
$\mathcal{M}_{1009}$			•	•	•	•	•	•	0.0180	0.0834
$\mathcal{M}_{645}$		•				•		•	0.0180	0.0834
$\mathcal{M}_{647}$	•	•				•		•	0.0161	0.0746
Empirical prior										
$\mathcal{M}_{915}$	•		•			•	•	•	0.0695	0.3161
$\mathcal{M}_{645}$		•				•		•	0.0495	0.2251
$\mathcal{M}_{661}$		•	•			•		•	0.0363	0.1651
$\mathcal{M}_{647}$	•	•				•		•	0.0332	0.1510
$\mathcal{M}_{899}$	•					•	•	•	0.0314	0.1428

Table 6.13: Models selected based on results from Gibbs variable selection for Bayesian model averaging. A bullet (•) indicates presence of a covariate in the model. The last column shows normalised posterior model probability.

To average the predicted consistency $\widetilde{w}_{1025}$ for patient 1025, $\widetilde{w}_{1025}^{(A)}$, we have:

$$\begin{aligned}\widetilde{w}_{1025}^{(A)} &= (0.1510)\widetilde{w}_{1025}^{(661)} + (0.1227)\widetilde{w}_{1025}^{(725)} + (0.1098)\widetilde{w}_{1025}^{(709)} + (0.1098)\widetilde{w}_{1025}^{(977)} \\ &\quad + (0.0908)\widetilde{w}_{1025}^{(753)} + (0.0889)\widetilde{w}_{1025}^{(663)} + (0.0857)\widetilde{w}_{1025}^{(737)} + (0.0834)\widetilde{w}_{1025}^{(1009)} \\ &\quad + (0.0834)\widetilde{w}_{1025}^{(645)} + (0.0746)\widetilde{w}_{1025}^{(647)}.\end{aligned}$$

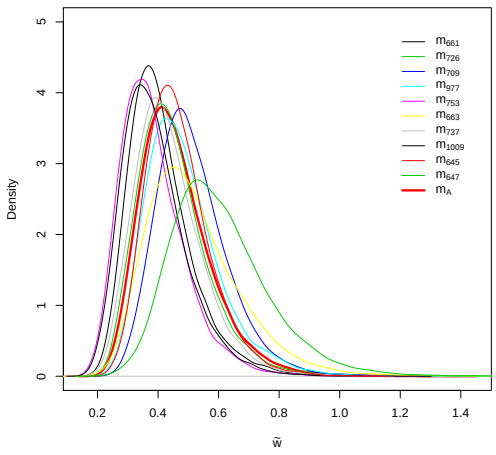
The posterior densities of the consistency parameter $\widetilde{w}_i$ from all ten models are given in Figure 6.1, and the red plot is the estimate with the averaged model. Results for subjects 1039, 3016, 5026, 6002, and 6064 are also shown in Figure 6.1.

Judging by the posterior odds, Gibbs variable selection with empirical prior (using $c^2 = n$) yields 5 believable models, as listed in Table 6.13. Model $\mathcal{M}_{915}$ is slightly favoured than model $\mathcal{M}_{645}$, with it being 1.4 times better than $\mathcal{M}_{645}$ (Table 6.11). These two models were averaged together with model $\mathcal{M}_{661}$, $\mathcal{M}_{647}$ and $\mathcal{M}_{899}$. The posterior distributions of the weighted average $\widetilde{w}_i^{(A)}$ obtained for six selected patients are given in Figure 6.2. For comparison purposes, posterior distributions of $\widetilde{w}_i^{(k)}$ for the other five potential models are plotted on the same axes.

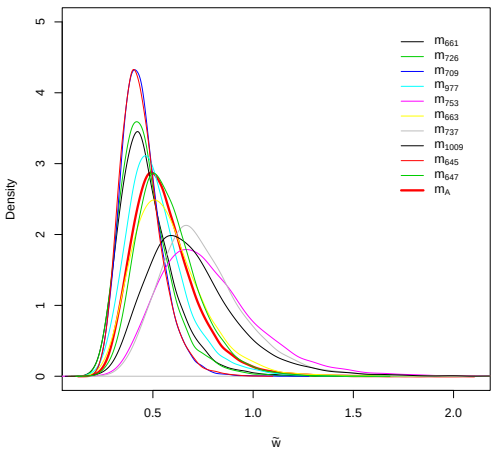
Only the top two models indicated by Gibbs variable selection using independent prior are selected for model averaging. The posterior distributions for $\widetilde{w}_i$ for six subjects based on the model averaging results are illustrated in Figure 6.3.

Some of the plots (see Figures 6.1 (b), 6.2 (c), and 6.3 (c)), show models that seem equally good (because their posterior odds are ≤ 3) but have considerably different $\widetilde{w}_i$ distributions from one another which may lead to different conclusion or prediction. This strongly shows the advantage of Bayesian model averaging where this method allows for the incorporation of model uncertainty into making inference based on a weighted average.

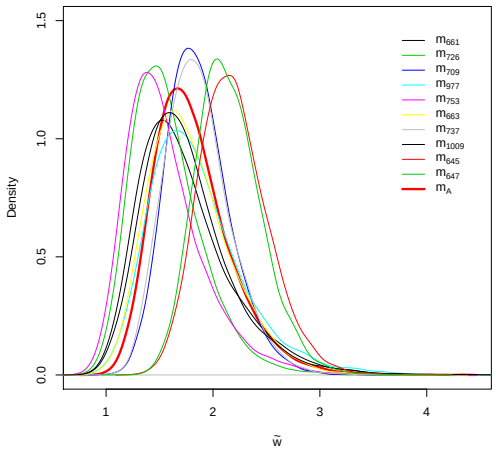
If we observe the results from Zellner's g-prior, covariate awareness of hypoglycaemia (x_7) is not included in the 'best' selected model which is the model with highest posterior probability. However, this covariate's probability of inclusion is 0.6198 indicating that this covariate may have predictive value and can be useful for prediction. By averaging the subset of models, this covariate is utilised in making prediction.



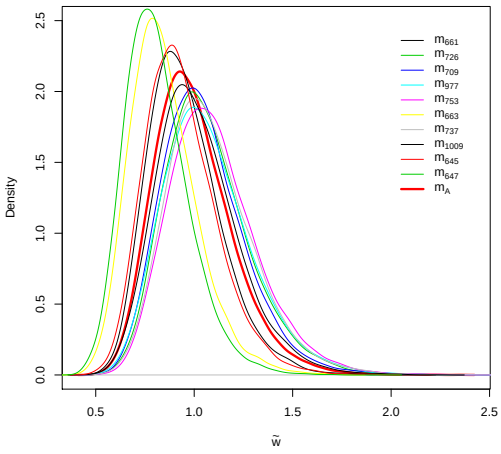
(a) Patient 1025



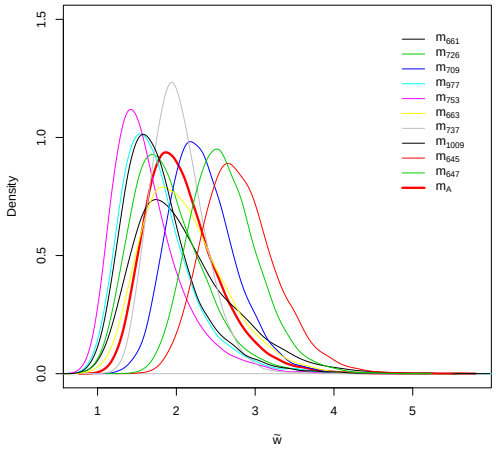
(b) Patient 1039



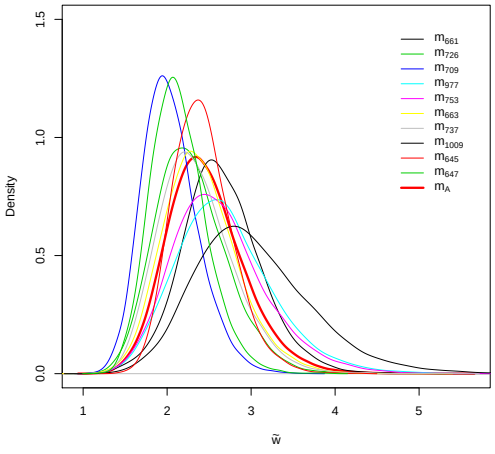
(c) Patient 3016



(d) Patient 5026

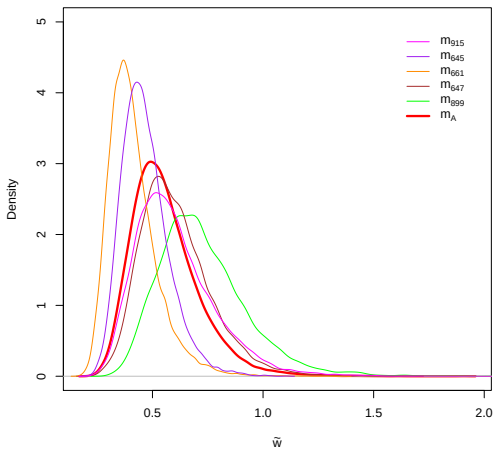


(e) Patient 6002

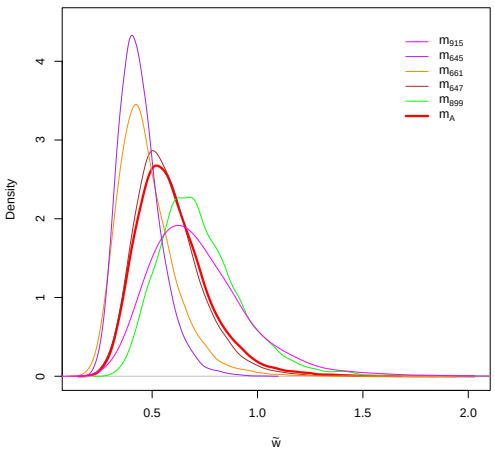


(f) Patient 6064

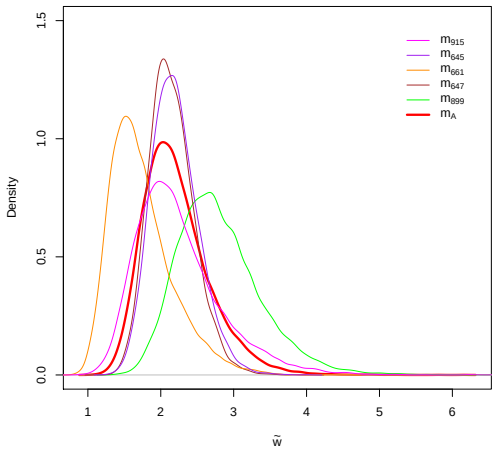
Figure 6.1: Posterior density plots of $\tilde{w}_i = \sigma_i^{-2}$ for different patients under selected models from Gibbs variable selection using Zellner's g-prior. The red plot represents the averaged model.



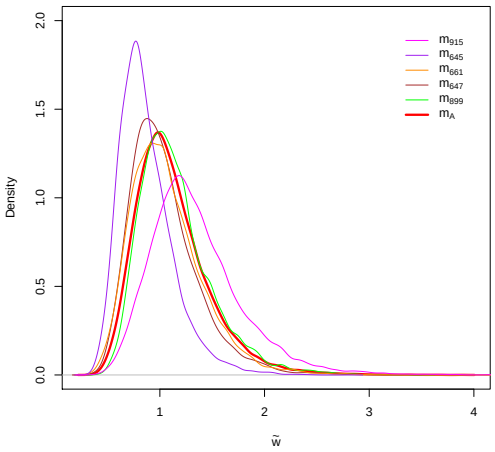
(a) Patient 1025



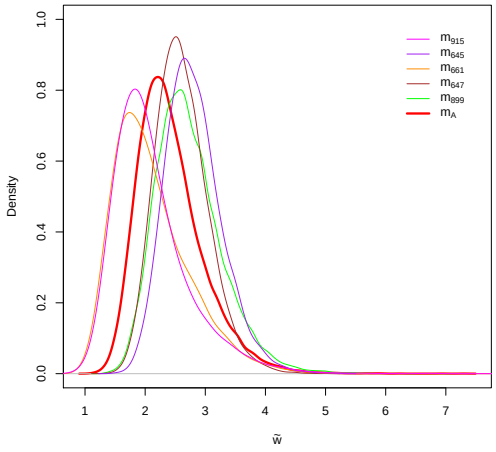
(b) Patient 1039



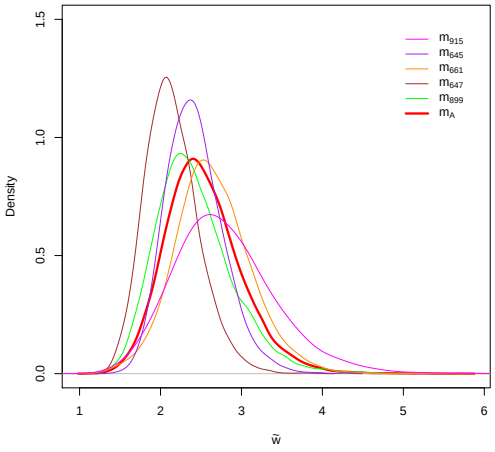
(c) Patient 3016



(d) Patient 5026

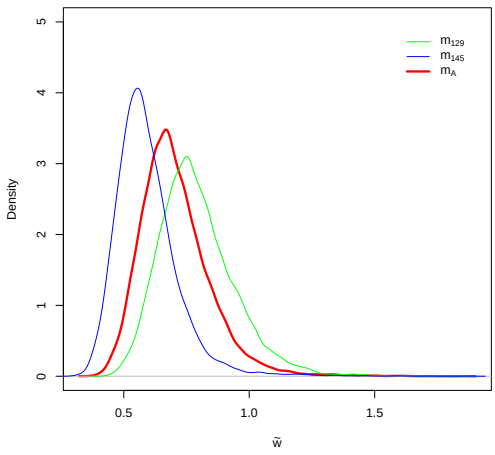


(e) Patient 6002

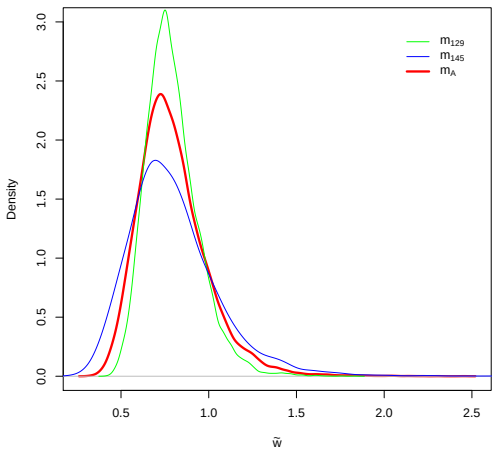


(f) Patient 6064

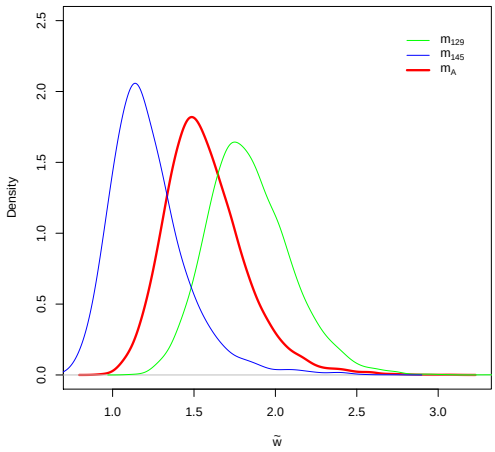
Figure 6.2: Posterior density plots of $\tilde{w}_i = \sigma_i^{-2}$ for different patients under selected models from Gibbs variable selection using empirical prior. The red plot represents the averaged model.



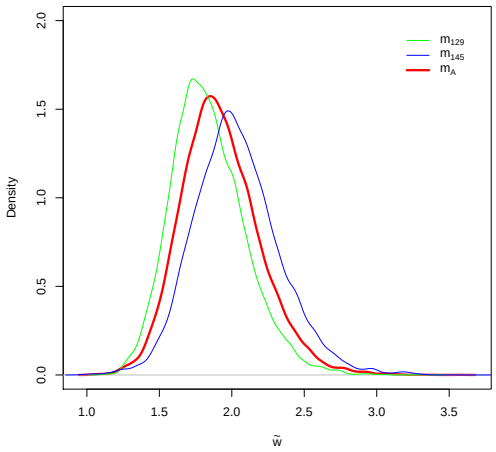
(a) Patient 1025



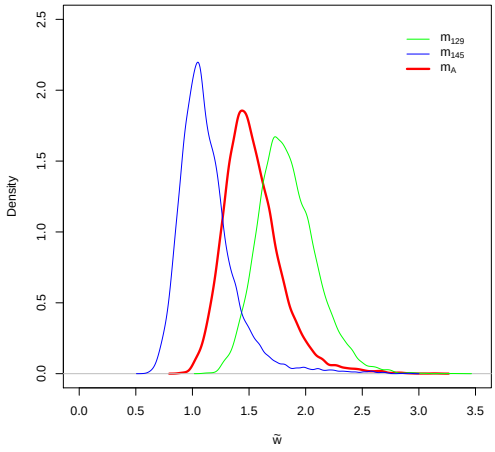
(b) Patient 1039



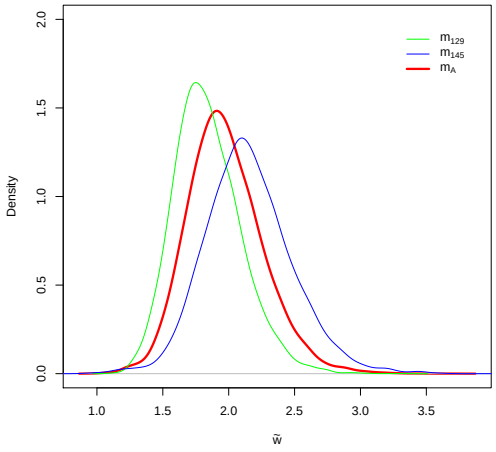
(c) Patient 3016



(d) Patient 5026



(e) Patient 6002



(f) Patient 6064

Figure 6.3: Posterior density plots of $\tilde{w}_i = \sigma_i^{-2}$ for different patients under selected models from Gibbs variable selection using independent prior. Red plot represents the averaged model.

6.4 Concluding remarks

In this chapter we discuss how we can select models that predicts patients' consistency well. We explore two methods which lead to different predictive models. We applied these methods to three models; the grouped symptoms model, the grouped symptoms model with interaction between covariates, and the hierarchical model with grouped symptoms. With the stepwise selection procedure, we end up with different models when starting the procedure from a different point (Tables 6.4 and 6.6). However, there is a number of common covariates chosen between the models. This indicates uncertainty that can lead to poor predictive ability.

Another method used to select the best predictive model is the Gibbs variable selection (GVS) where we consider three prior distribution settings for the coefficient $\boldsymbol{b}$. Based on the results discussed in Section 6.2, the best model when using an independent prior is $\mathcal{M}_{129}$ which contains covariate gender, while with Zellners g-prior results are dependent on the prior variance. With larger c^2 the results are more consistent. Although $\mathcal{M}_{641}$ (model with covariates gender and retinopathy) is the best model when $c^2 = 1000$, this model comes second when using $c^2 = 10{,}000$ with posterior odds 0.63. On the other hand, with the empirical prior the best model is the model containing covariates duration of diabetes, Hba, gender, type of diabetes and retinopathy ($\mathcal{M}_{915}$). Even though we get different best models with different prior settings, we can conclude that they are not inconsistent since models selected with posterior odds ≤ 3 in each prior setting are similar.

We then counter the uncertainty issue shown in the GVS method by using Bayesian model averaging methodology. We select a subset of models obtained from GVS, which consists of models with posterior odds ≤ 3, and average them for making predictions. However, if there are more than ten models with posterior odds ≤ 3 we limit the averaging to just ten models as the posterior probability for other models can be very small.

Chapter 7

Conclusions and Further Research

7.1 Conclusions

The main contributions of this thesis are the development and assessment of statistical logistic-type models for the consistency of symptoms experienced by individual patients during hypoglycaemia using MCMC methodology under a Bayesian approach.

We started by exploring different options for the symptom experiencing threshold $h(\alpha_{ij}, \beta_{ik})$ to fit to the core consistency estimation model. Model selection for the best threshold was conducted by considering both the deviance information criterion and likelihood comparisons for models with different thresholds. This comparison suggested that the most suitable threshold for this model is $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$. The sensitivity of our estimation to prior assumptions was also investigated, and we demonstrated that our choice of priors did not have an important impact on the ranking of the consistency estimates for the 66 patients considered. We then expanded the consistency estimation model by incorporating between-group variability, when symptoms of hypoglycaemia are classified into six different groups depending on the cause of symptoms. The consistency of symptoms experienced by individual patients with this calibrated model, is similar to that obtained with an earlier model (Zammitt et al., 2011) suggesting that individuals show variation in consistency. Regarding the grouped symptoms, autonomic and neuroglycopenic groups have the highest propensity, implying that symptoms categorised in these groups are important in detecting the onset of hypoglycaemia.

Model assessment was performed to explore the adequacy of the grouped symptoms models using the concept of stochastic latent residuals. This method was implemented to the model with thresholds $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ and $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij} + \beta_{ik}$. Preliminary checking showed that residuals for each patient come from a Uniform(0,1) distribution, indicating a good fit. Furthermore, from the posterior distribution of p-values, π_γ, obtained with a Kolmogorov-Smirnov goodness-of-fit test, we concluded that there is no strong evidence against the model tested. We used $\Pr(\pi_\gamma < 0.05)$ as a measure of fit and it was found that the model with grouped symptoms using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$ fits the data better.

We verified the predictive ability of our model by comparing the observed data with predicted values using an approach based on the posterior predictive distribution. The model with grouped symptoms using threshold $h(\alpha_{ij}, \beta_{ik}) = \alpha_{ij}\beta_{ik}$, gave posterior distribution of predicted number of reported symptoms containing the observed number, implying that this model performs well in predicting consistency.

Earlier work was restricted to only using episodes not occurring within 24 hours, as these episodes were thought to have the potential to reduce the intensity of following episodes. In Chapter 4, we examined this claim by examining the correlation and level of intensity between episodes occurring within 24 hours and episodes not within 24 hours. There is no considerable difference between the correlation of intensity and the average intensity between the two cases. Due to these results, we concluded that there is no evidence of episodes occurring within 24 hours having a considerable effect on intensity. Following that, we also used permutation testing to examine if episodes' intensity shows association between cases where there are episodes within 24 hours and cases where those episodes are excluded. Again, results showed no evidence against the hypothesis of no association. Therefore, subsequent analysis in the thesis included data from all episodes regardless of their time of occurrence. Episodes with missing date were also taken into account since it is not necessary to determine their time of occurrence. Analysis with all episodes gave higher patients' consistency as compared to when excluding episodes within 24 hours. Also, more reliable estimates

are obtained when all episodes are used. This was indicated by lower coefficient of variation values when we used the additional episodes. Overall, the distribution of the consistency estimates does not significantly change after the inclusion of the 24 hours episodes.

It is crucial for diabetic patients to understand and be aware of which factors may cause variation in their symptoms of hypoglycaemia. Using generalised linear model methodology and posterior precision estimates, $\tilde{\sigma}_i^{-2}$, under the model with grouped symptoms, it was found that female patients are less consistent in reporting symptoms of hypoglycaemia, which agrees with previous findings. Our results also suggested that another significant factor affecting consistency is lack of retinopathy. We have also examined the impact of interactions between covariates on consistency. Eleven interactions were considered. According to the estimation results, modelling with interactions between covariates agrees with our earlier findings where gender and 'no retinopathy' appear to be significant. We also conclude that female patients experience reduction in consistency as their duration of diabetes and BMI measure increase. When compared to male patients, female subjects with no retinopathy show lower consistency, but those with proliferative retinopathy have higher consistency. For patients with no retinopathy, their consistency is higher if they have longer duration of diagnosed diabetes. Besides that, the interaction between duration and C-peptide level also significantly affects consistency, such that patients with long duration of diabetes and high level of C-peptide have lower consistency estimates compared to patients with shorter duration and lower C-peptide level.

We have also developed a hierarchical model that is able to estimate consistency and determine factors affecting it in a single model. This is done by giving a hierarchical prior to the variance parameter, σ_i^2. Similar to the conclusion made with the non-hierarchical model, this extended consistency model shows that consistency of symptoms reporting among individual varies. However, when the hierarchical setting is introduced the strong negative effect of gender disappears. In fact, no factor is found significant.

The thesis also includes work on variable selection to determine the most suitable model to predict patients' consistency using the model with grouped symptoms. We used stepwise selection and Gibbs variable selection methodologies with different prior settings for the model parameters. In stepwise selection, different initial points were used resulting in two final 'best models'. Starting the selection process with full model, a model excluding age and duration was selected as the best predictive model. However, with the null model and model containing covariate gender as the starting model, a final model excluding age, duration, C-peptide and Ace was supported. We repeated the same procedure including interactions between covariates. As can be seen in Table 6.6, different starting points favour different predictor models. However, the two resulting models have selected six common predictors. Applying the same methodology to the hierarchical model gives a more consistent result, as regardless of the starting points the selected predictive model is the same; this is the model with covariates C-peptide, Ace, gender, type of diabetes, and retinopathy.

The second approach to variable selection was using Gibbs variable selection with an independent prior, Zellner's g-prior, and an empirical prior. The model with covariate gender is found to be the best with the independent prior, but when using Zellner's g-prior a larger set of covariates is included in the predictive model. However, a more parsimonious model is obtained as the prior variance increases. With the empirical prior setting the best determined model is that containing covariates duration, Hba, gender, type of diabetes, and retinopathy. Finally, we overcame the issue of uncertainty in Gibbs variable selection with Bayesian model averaging.

Finally, we stress again here that as the sample of patients used in our analyses is not necessarily representative of the population of diabetic patients suffering hypoglycaemic episodes, we do not consider our results to be directly generalised to this entire population. However, our findings can be used to guide and inform studies in patient groups with similar characteristics, and our methodology can be easily applied to expanded data sets (where available) to provide more general conclusions.

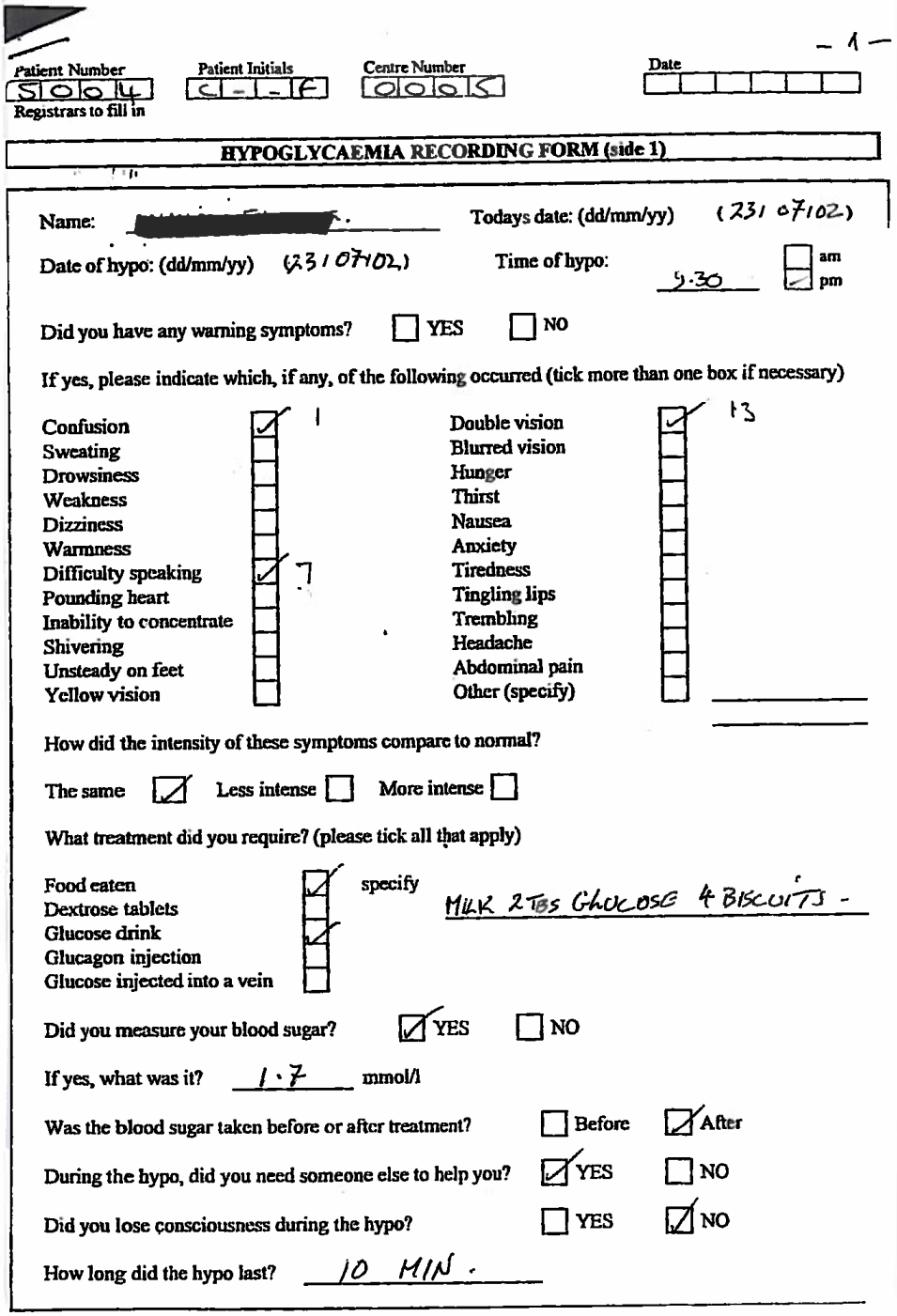
7.2 Further research

An interesting topic for future research would be exploring other distributions for the threshold variable τ_{ijk}. Among potential distributions are gamma, Weibull, and Burr. Other positive-value distributions can also be considered. A different extension to this thesis can be provided by modelling the correlation between episode intensity through introducing correlation in the prior distributions of the parameters β_{ik}, which represent to the intensity of the episodes of hypoglycaemia. It is possible to consider using conditional autoregressive priors (Carlin and Banerjee, 2003) which mainly model the dependence between neighbouring episodes. Such a model could be useful for further determining the correlation between episodes' intensity, especially for cases which occur within 24 hours.

Finally, further research can be performed with new or follow-up data in future. The models and methodology developed in this thesis can be adapted and extended to apply to richer data, in order for example, to investigate stronger association of symptom consistency with patient-specific covariates. Also, more informative data might help to improve the power of assessment and model diagnostics presented in this thesis, and could also help distinguish better between competing models.

Appendix A

Sample of Report Form



- 1 -

Patient Number: 5004 | Patient Initials: [redacted] | Centre Number: 0005 | Date:

Registrars to fill in

HYPOGLYCAEMIA RECORDING FORM (side 1)

Name: [redacted] Todays date: (dd/mm/yy) (23107102)

Date of hypo: (dd/mm/yy) (23107102) Time of hypo: 5.30 ☐ am ☑ pm

Did you have any warning symptoms? ☐ YES ☐ NO

If yes, please indicate which, if any, of the following occurred (tick more than one box if necessary)

Confusion	☑ 1	Double vision	☑ 13
Sweating	☐	Blurred vision	☐
Drowsiness	☐	Hunger	☐
Weakness	☐	Thirst	☐
Dizziness	☐	Nausea	☐
Warmness	☐	Anxiety	☐
Difficulty speaking	☑ 7	Tiredness	☐
Pounding heart	☐	Tingling lips	☐
Inability to concentrate	☐	Trembling	☐
Shivering	☐	Headache	☐
Unsteady on feet	☐	Abdominal pain	☐
Yellow vision	☐	Other (specify)	☐ _____

How did the intensity of these symptoms compare to normal?

The same ☑ Less intense ☐ More intense ☐

What treatment did you require? (please tick all that apply)

Food eaten	☑	specify MILK 2 TBS GLUCOSE 4 BISCUITS -
Dextrose tablets	☐	
Glucose drink	☑	
Glucagon injection	☐	
Glucose injected into a vein	☐	

Did you measure your blood sugar? ☑ YES ☐ NO

If yes, what was it? 1.7 mmol/l

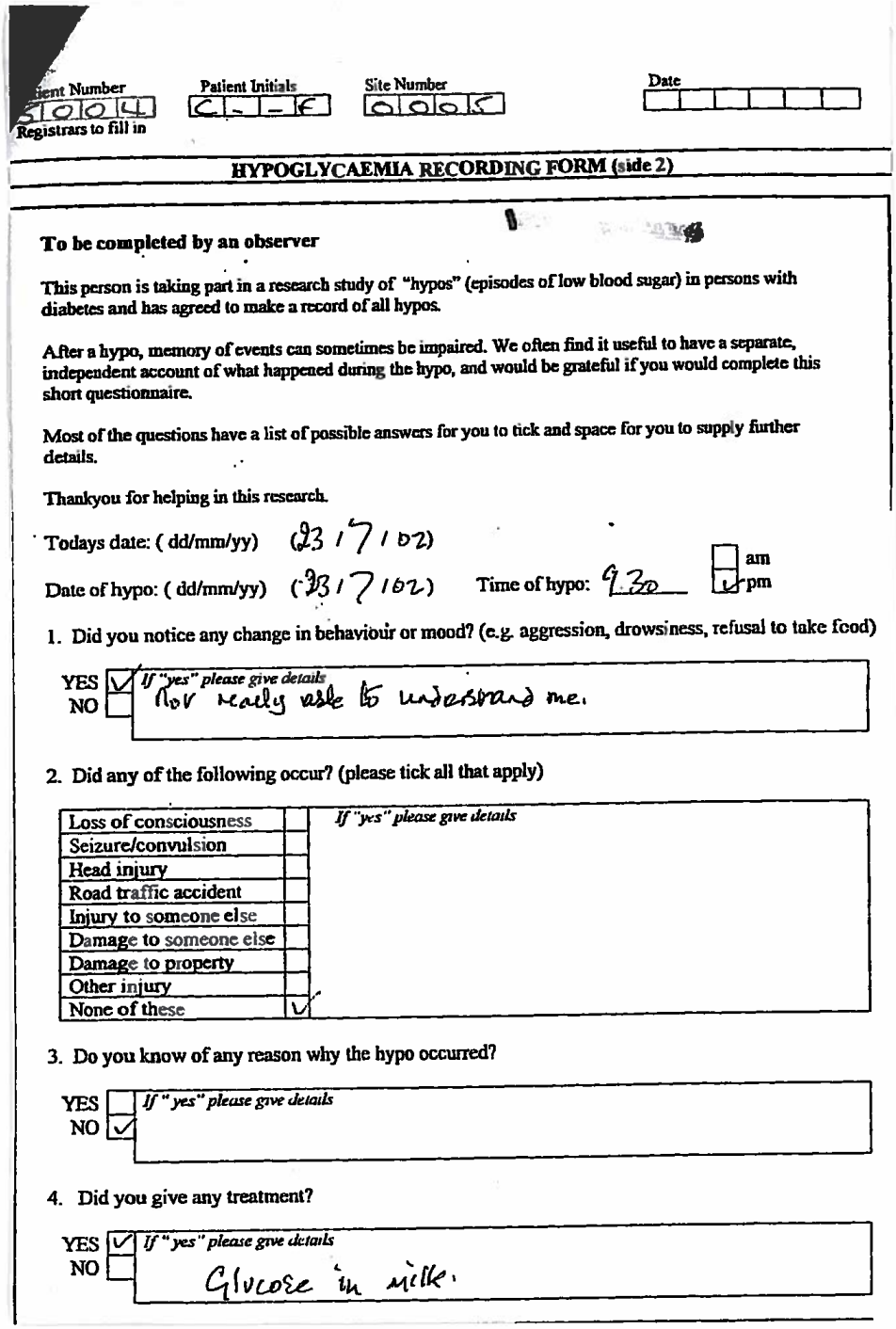
Was the blood sugar taken before or after treatment? ☐ Before ☑ After

During the hypo, did you need someone else to help you? ☑ YES ☐ NO

Did you lose consciousness during the hypo? ☐ YES ☑ NO

How long did the hypo last? 10 MIN.

Figure A.1: Sample of report form (page 1).



ient Number 5 0 0 0 4 Registrars to fill in

Patient Initials C - - F

Site Number 0 0 0 5

Date

HYPOGLYCAEMIA RECORDING FORM (side 2)

To be completed by an observer

This person is taking part in a research study of "hypos" (episodes of low blood sugar) in persons with diabetes and has agreed to make a record of all hypos.

After a hypo, memory of events can sometimes be impaired. We often find it useful to have a separate, independent account of what happened during the hypo, and would be grateful if you would complete this short questionnaire.

Most of the questions have a list of possible answers for you to tick and space for you to supply further details.

Thankyou for helping in this research.

Todays date: (dd/mm/yy) (23/7/02)

Date of hypo: (dd/mm/yy) (23/7/02) Time of hypo: 9.30 ☐ am ☑ pm

1. Did you notice any change in behaviour or mood? (e.g. aggression, drowsiness, refusal to take food)

YES ☑ NO ☐ *If "yes" please give details* Not really able to understand me.

2. Did any of the following occur? (please tick all that apply)

Loss of consciousness	☐	*If "yes" please give details*
Seizure/convulsion	☐	
Head injury	☐	
Road traffic accident	☐	
Injury to someone else	☐	
Damage to someone else	☐	
Damage to property	☐	
Other injury	☐	
None of these	☑	

3. Do you know of any reason why the hypo occurred?

YES ☐ NO ☑ *If "yes" please give details*

4. Did you give any treatment?

YES ☑ NO ☐ *If "yes" please give details* Glucose in milk.

Figure A.2: Sample of report form (page 2).

Appendix B

Covariates for All Patients

Subject	GEN	AGE	TYPE	DUR	RET	BMI	CPEP	HBA	ACE	AWAR
1008	M	67	1	306	2	21.6	0.1	6.7	28	3
1009	M	41	1	60	1	27.9	0.18	7.6	43	3
1015	F	34	1	204	2	42.7	0.23	8.5	44	1
1021	M	39	1	36	1	24.9	0.13	9	24	3
1025	F	49	1	540	2	25.8	0.05	8.6	68	3
1028	M	51	1	324	1	29.7	0.05	NA	28	7
1036	M	31	1	12	1	21.2	0.62	10.1	33	2
1039	F	70	1	540	2	24.8	0.2	NA	5	2
1055	M	65	2	84	2	25.6	0.07	7	49	3
1057	F	43	2	120	1	37.5	0.69	6.5	38	1
1086	F	55	1	336	2	37.5	0.05	8.3	10	1
2009	F	57	1	540	3	33	0.22	7.2	30	5
2010	F	59	1	636	1	25.7	0.05	6.6	3	6
2012	M	34	1	48	1	27.4	0.45	6.8	63	2
2013	F	60	1	576	3	26.4	0.42	7	43	3
2015	F	58	1	432	2	22.1	0.14	7.8	33	6
2021	M	50	1	408	2	24.6	0.05	NA	4	2
2022	M	38	1	444	2	24.5	0.83	7.7	48	2
2027	M	26	1	57	1	22.5	0.62	7.9	28	6
3001	M	66	1	12	1	22	0.69	7.6	21	1
3015	F	60	1	180	1	22	0.05	7.2	98	3
3016	M	65	2	84	1	26	0.27	8.5	24	1
3022	M	58	1	300	2	32.4	0.05	7.1	7	1
3024	F	52	1	48	1	21	0.34	5.6	41	1
3029	M	60	1	36	1	23.8	0.33	6.5	18	1
3043	M	62	2	36	1	37	1.21	6.3	12	1
3046	M	51	2	24	2	23.7	2.22	7.1	20	1
3048	M	68	2	300	2	33	0.09	6.3	14	1
3050	M	49	1	36	1	25	0.06	NA	94	3
3052	M	57	2	108	1	25	0.24	8.4	8	2
3057	M	63	2	60	1	27	0.05	8.1	39	2
3065	F	68	2	408	NA	24.5	1.13	7.6	63	2
3067	M	61	2	372	3	33	2.51	7.8	4	1
4003	M	50	1	408	2	31.8	0.08	8.8	6	7
4008	M	62	1	528	3	23.9	0.05	8.3	50	5
4013	F	54	1	360	2	24.5	0.84	6.5	43	1
4023	M	46	1	57	1	23	0.59	NA	38	1
4028	M	69	2	168	1	25.1	0.97	7.5	46	1
4032	M	64	2	204	1	25.7	1.31	7.2	7	1
4034	M	33	1	30	1	23.8	NA	6.3	55	2
4043	M	61	1	288	2	21.9	0.85	6.5	9	3
4045	F	63	1	552	3	23.1	0.05	8.3	63	1
4049	F	22	1	6	1	19.5	0.34	7.3	50	1
4061	F	71	2	60	1	27.3	0.33	9.5	20	1
4063	M	70	1	18	2	27.7	0.64	7.2	76	3
4072	M	68	2	120	3	21.9	0.08	6.9	71	5
4076	M	58	2	156	1	28	0.52	7.5	6	1
5004	M	67	1	252	2	26.1	0.23	8	57	3
5009	M	72	2	108	3	30.5	1.69	8.9	41	1
5023	M	72	1	744	3	28.6	0.31	9	7	4
5026	M	69	1	540	3	22.7	0.05	6.6	24	1
5029	F	70	1	12	1	24.3	0.38	6.2	25	1
5044	F	22	1	36	1	26	0.14	6.8	22	2
5045	M	38	1	30	1	29.6	0.25	7.7	22	2
5088	F	68	1	18	1	24	0.82	7.2	50	3
6002	M	65	2	72	1	30.2	1.24	8.8	16	1
6010	M	42	1	36	1	23.3	0.28	6.9	34	1
6018	F	44	1	384	2	26	0.14	7.1	41	1
6019	F	36	1	24	1	20.5	0.87	NA	29	5
6023	F	51	1	444	1	23.5	0.07	6.1	41	6
6038	M	32	1	24	1	25.2	0.54	6.6	34	1
6056	M	74	2	12	1	26.8	0.45	8.1	7	1
6058	M	60	2	108	1	28.8	1.58	7.8	11	1
6062	M	61	2	18	1	27.7	1.12	6.7	23	1
6064	M	59	2	12	1	27.7	0.8	6.4	2	1
6065	F	34	1	36	1	24.3	0.61	7.7	32	1

Table B.1: List of covariates for all patients where RET 1 = no retinopathy; RET 2 = background retinopathy and RET 3 = proliferative retinopathy; hypoglycaemia awareness (AWAR) score is from 1 to 7, with higher scores corresponding to weaker awareness of hypoglycaemia. See Section 5.1 for full details.

Bibliography

Arminger, G. and Muthen, B. O. (1998). A bayesian approach to nonlinear latent variable models using the gibbs sampler and the metropolis-hastings algorithm. *Psychometrika*, 63(3):271–300.

Association, A. D. (1966). *Diabetes*, volume 15. American Diabetes Association.

Association, A. D. et al. (2005). *American Diabetes Association Clinical Practice Recommendations 2005*. American Diabetes Association.

Atkinson, M. A. and Maclaren, N. K. (1994). The pathogenesis of insulin-dependent diabetes mellitus. *New England Journal of Medicine*, 331(21):1428–1436. PMID: 7969282.

Banarer, S. and Cryer, P. E. (2004). Hypoglycemia in type 2 diabetes. *Medical Clinics of North America*, 88(4):1107–1116.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.

Blair, R. and Karniski, W. (1994). Distribution-free statistical analyses of surface and volumetric maps. *Functional Neuroimaging: Technical Foundations*.

Bogardus, C., Lillioja, S., Mott, D., Hollenbeck, C., and Reaven, G. (1985). Relationship between degree of obesity and in vivo insulin action in man. *American Journal of Physiology-Endocrinology And Metabolism*, 248(3):E286–E291.

Carlin, B. P. and Banerjee, S. (2003). Hierarchical multivariate car models for spatio-temporally correlated survival data. *Bayesian statistics*, 7:45–63.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–673.

Committee, C. D. A. C. P. G. E. (2003). *Canadian Diabetes Association 2003 clinical practice guidelines for the prevention and management of diabetes in Canada*. Can J Diabetes.

Cox, D. J., Gonder-Frederick, L., Antoun, B., Cryer, P. E., and Clarke, W. L. (1993). Perceived symptoms in the recognition of hypoglycemia. *Diabetes Care*, 16(2):519–527.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 248–275.

Dawid, A. P. and Stone, M. (1982). The functional-model basis of fiducial inference. *The Annals of Statistics*, pages 1054–1067.

Deary, I., Hepburn, D., MacLeod, K., and Frier, B. (1993). Partitioning the symptoms of hypoglycaemia using multi-sample confirmatory factor analysis. *Diabetologia*, 36(8):771–777.

Dellaportas, P., Forster, J. J., and Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12:27–36.

Edgington, E. S. (1987). Randomization tests. 2nd ed.

Fisher, R. (1971). Statistical methods, experimental design, and scientific inference.

Fisher, R. A. (1992). *On the Mathematical Foundations of Theoretical Statistics*, pages 11–44. Springer New York, New York, NY.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. (1990). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). Bayesian data analysis.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 6(4):733–760.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741.

Group, N. D. D. (1979). Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes*, 28(12):1039–1057.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Henderson, J., Allen, K., Deary, I., and Frier, B. (2003). Hypoglycaemia in insulin-treated type 2 diabetes: frequency, symptoms and impaired awareness. *Diabetic Medicine*, 20(12):1016–1021.

Hepburn, D., Deary, I., and Frier, B. (1992). Classification of symptoms of hypoglycaemia in insulin-treated diabetic patients using factor analysis: Relationship to hypoglycaemia unawareness. *Diabetic Medicine*, 9(1):70–75.

Hepburn, D. A., Deary, I. J., Frier, B. M., Patrick, A. W., Quinn, J. D., and Fisher, B. M. (1991). Symptoms of acute insulin-induced hypoglycemia in humans with and without iddm: factor-analysis approach. *Diabetes Care*, 14(11):949–957.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2002). Bayesian methods for generalized linear models with covariates missing at random. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 55–78.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186(1007):453–461.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

Kolterman, O., Insel, J., Saekow, M., and Olefsky, J. (1980). Mechanisms of insulin resistance in human obesity: evidence for receptor and postreceptor defects. *Journal of Clinical Investigation*, 65(6):1272.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89(428):1535–1546.

Manly, B. F. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. CRC Press.

McAulay, V., Deary, I., and Frier, B. (2001). Symptoms of hypoglycaemia in people with diabetes. *Diabetic Medicine*, 18(9):690–705.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Miller, A. (2002). *Subset Selection in Regression*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

Nattrass, M. and Lauritzen, T. (2000). Review of prandial glucose regulation with repaglinide: a solution to the problem of hypoglycaemia in the treatment of type 2 diabetes? *International Journal of Obesity*, 24:S21–S31.

Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS*, volume 698. Wiley.

Ntzoufras, I., Dellaportas, P., and Forster, J. J. (2003). Bayesian variable and link determination for generalised linear models. *Journal of statistical planning and inference*, 111(1):165–180.

Ntzoufras, I. et al. (2002). Gibbs variable selection using bugs. *Journal of Statistical Software*, 7(7):1–19.

Olefsky, J. M., Kolterman, O. G., and Scarlett, J. A. (1982). Insulin action and resistance in obesity and noninsulin-dependent type ii diabetes mellitus. *American Journal of Physiology-Endocrinology And Metabolism*, 243(1):E15–E30.

R Core, T. (2015). R: A language and environment for statistical computing [internet]. vienna, austria: R foundation for statistical computing; 2013. *Document freely available on the internet at: http://www. r-project. org.*

Raab, G. (2014). Gillian raabs materials on winbugs. `http://www-users.york.ac.uk/~pml1/bayes/winbugsinfo/raab.htm`. Accessed: 2016-01-11.

Raab, G., Yang, S., Allardice, G., Goldberg, D., and McMenamin, J. (1998). Modelling human immunodeficiency virus infection and acquired immune deficiency syndrome cases in scotland: data sources, prior information and bayesian estimation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161(3):367–384.

Reaven, G. M., Bernstein, R., Davis, B., and Olefsky, J. M. (1976). Nonketotic diabetes mellitus: insulin deficiency or insulin resistance? *The American journal of medicine*, 60(1):80–88.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, pages 1151–1172.

Schlaifer, R. and Raiffa, H. (1961). Applied statistical decision theory.

Sotiropoulos, A., Skliros, E., Tountas, C., Apostolou, U., Peppas, T., and Pappas, S. (2005). Risk factors for severe hypoglycaemia in type 2 diabetic patients admitted to hospital in piraeus, greece.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). Winbugs user manual. *Cambridge: MRC Biostatistics Unit.*

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Streftaris, G. and Gibson, G. (2012). Non-exponential tolerance to infection in epidemic systems - modeling, inference, and assessment. *Biostatistics*, 13(4):580–593.

Streftaris, G., Wallerstein, N., Gibson, G., and Arthur, S. (2013). Modeling probability of blockage at culvert trash screens using bayesian approach. *Journal of Hydraulic Engineering*, 139(7):716–726.

Towler, D. A., Havlin, C. E., Craft, S., and Cryer, P. (1993). Mechanism of awareness of hypoglycemia: perception of neurogenic (predominantly cholinergic) rather than neuroglycopenic symptoms. *Diabetes*, 42(12):1791–1798.

Weisberg, S. (1985). Applied linear regression, a john wiley & sons. *New York*, 324.

Xie, Y. and Carlin, B. P. (2006). Measures of bayesian learning and identifiability in hierarchical models. *Journal of Statistical Planning and Inference*, 136(10):3458–3477.

Zammitt, N., Streftaris, G., Gibson, G., Deary, I., and Frier, B. (2011). Modeling the consistency of hypoglycemic symptoms: high variability in diabetes. *Diabetes Technology & Therapeutics*, 13(5):571–578.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243.